Theoretical Design and Computational Analysis of Environmental Effects on Carbon Bowl Alkylated Annelated Corannulene

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Yohann Potier

aus Frankreich

<u>Promotionskomitee</u>:

Prof. Dr. Kim K. Baldridge (Vorsitz)

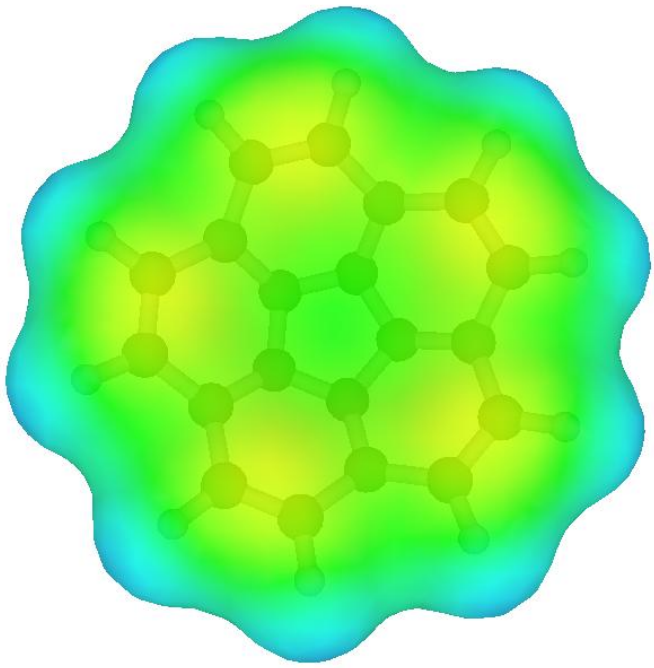
Prof. Dr. Jay S. Siegel

Prof. Dr. Reto Dorta

Dr. Karl-Heinz Ernst

Zürich, 2009

Theoretical Design and Computational Analysis of Environmental Effects on Carbon Bowl Alkylated Annelated Corannulene



ABSTRACT OF THE DISSERTATION

Theoretical Design and Computational Analysis of Environmental Effects on Carbon Bowl Alkylated Annelated Corannulene

by

Yohann Potier

University of Zürich, 2009

Professor Kim K. Baldridge, Chair

The presented work focuses on the development and application of quantum chemical based methodology for prediction of structure and properties of molecules in an environment. In this context, consideration of gas phase, solution phase, and crystalline environments are considered.

For consideration of solvent environment, the developed methodology involves a highly accurate implicit solvation model embedded in the quantum chemical modeling program, GAMESS (General Atomic Molecular Electronic Structure Systems). Implementations of key new features of the methodology include a) cavity construction improvement b) a better description of the charge distribution and c) inclusion of non-electrostatic effects. The model is discussed in detail, validated, and illustrated through examples associated with important structural, energetic and property effects on key organic systems. An important goal includes careful consideration of issues associated with first solvation shell effects. Through the use of convenient statistical mechanics-based treatments, very accurate solvation properties, such as solvation energies, pKa, and polarization charge density, can be illustrated.

For consideration of crystal environment, the work involves exploitation of a quantum mechanical molecular electron density representation to derive optimal crystal packing descriptions and corresponding intermolecular energies based on empirical terms. These studies provide experimentalist with a framework to better understand experimental phenomenon, including exhibited properties.

These methods are validated for well-known crystal structures, and subsequently used for prediction in collaborative research involving polycyclic aromatic hydrocarbon systems.

Polycyclic aromatic hydrocarbon systems all stem from sheets of sp2 hybridized carbon arranged in aromatic/benzenoid subunits. Corannulene is a fundamental structure in this family. Corannulene is rich in structural, dynamic and physical features that challenge modern density functional and quantum mechanical methods. Applications of developed theoretical methods are applied to several topics concerning corannulene-based systems. In these topics, emphasis is placed on structure and dynamical behavior of various functionalized systems in the gas phase, solution phase, and crystal environment.

The thesis is broken down into i) background and motivation, ii) description of pertinent methodology, iii) quantum mechanical-based solvation implementations, iv) quantum mechanical-based crystal packing theory, v) research investigations, vi) conclusions and perspectives.

ZUSAMMENFASSUNG

Theoretical Design and Computational Analysis of Environmental Effects on Carbon Bowl Alkylated Annelated Corannulene

von

Yohann Potier

Universität Zürich, 2009

Kim K. Baldridge, Vorsitz

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung und Anwendung einer auf Quantenchemie basierenden Methode, Strukturen und Eigenschaften von Molekülen in einer bestimmten Umgebung vorherzusagen. In diesem Zusammenhang werden Gasphase, Lösung und der kristalline Zustand berücksichtigt.

Für die Untersuchung des flüssigen Zustandes wird eine Einbettung eines akkuraten impliziten Solvatationsmodells in das quantenchemische Modeling-programm GAMESS verwendet (General Atomic Molecular Electronic Structure Systems). Wichtige Eigenschaften dieser Methode sind a) eine verbesserte Beschreibung von Hohlräumen, b) eine verbesserte Beschreibung der Ladungsverteilung und c) der Einbezug von nicht-elektrostatischen Effekten. Das Modell wird detailliert erläutert, evaluiert und anhand von Effekten auf die Struktur, Energie und Eigenschaften ausgewählter organischer Verbindungen veranschaulicht. Ein prioritäres Ziel stellt die sorgfältige Untersuchung von Problemen dar, die im Zusammenhang stehen mit den Effekten durch die engste Solvatationshülle. Durch die Anwendung von Strategien aus der statistischen Mechanik können Solvatationseffekte wie beispielsweise Solvatationsenergien, pK_a-Werte und Polarisationsladungsdichten genau bestimmt werden.

Der kristalline Zustand wird mit Hilfe einer quantenmechanischen Darstellung der molekularen Elektronenverteilung beschrieben, so dass bevorzugte Kristallpackungen und entsprechende intermolekulare Energien, basierend auf empirischen Termen, vorhergesagt werden können. Diese Untersuchungen können Experimental-

chemikern eine Grundlage zum besseren Verständnis beobachteter Phänomene liefern. Die Methoden werden anhand wohlbekannter Kristallstrukturen geprüft und darüber hinaus in interdisziplinärer Forschung angewandt auf die Vorhersage der Festkörperstrukturen polyzyklischer aromatischer Kohlenwasserstoffe.

Polyzyklische aromatische Kohlenwasserstoffe bestehen aus annähernd planar angeordneten, sp^2-hybridisierten Kohlenstoffatomen. Corannulen ist ein wichtiger Vertreter dieser Klasse. Es ist ein Molekül, das reich ist an strukturellen, dynamischen und physikalischen Eigenschaften, deren theoretische Beschreibung eine Herausforderung für die moderne Dichtefunktionalstheorie und quantenmechanische Methoden darstellt. Theoretische Methoden werden auf verschiedene Eigenschaften Corannulen-basierter Systeme angewandt. Besonders die Struktur und das dynamische Verhalten in der Gasphase, in Lösung und im festen Zustand werden untersucht.

Die Dissertation ist gegliedert in i) Hintergrund und Motivation, ii) Beschreibung der verwendeten Methoden, iii) quantenmechanisch basierte Solvatationsmodelle, iv) quantenmechanisch basierte Kristallpackungstheorie, v) Forschungsergebnisse und vi) Schlussfolgerungen und Ausblick.

Curriculum Vitae

2004-2008:	**PhD. in Computational Chemistry** Organic Chemistry Institute - University of Zurich, Switzerland
2002-2003:	**Master of Biotechnology and Informatics** **DESS (Superior Specialized Studies Diploma) " Génie Physiologique et Informatique "** University of Poitiers, France – with honors
2001-2002:	**MA Degree in Physiology and Computer Engineering** University of Poitiers, France – with honors
2000-2001:	**MA Degree in Biology with Specialization in Animal Physiology** University of Rennes, France – with honors
1998-2000:	**BA Degree in Biology with Specialization in Physiology** University of Rennes, France
1995-1998:	**DEUG " Life and Earth Sciences " (First part of BA Degree in Biology)** University of Rennes, France
1995:	**French High School Diploma** Sévigné High School, France

Presentation and invited lectures

"Computational analysis of environment impact on Corannulene" Swiss Chemical Society Fall Meeting 2005, Lausanne

"Quantum mechanical solvation methods: GAMESS-COSab method" Dorothy Crowfoot Hodgkin Symposium 2006, Zurich

"Application of a Quantum Mechanical solvation method for theoretical Corannulene studies" Swiss Chemical Society Fall Meeting 2007, Lausanne

Publications

"The Computational Chemistry Prototyping Environment"
K. K. Baldridge, J. P. Greenberg, W. Sudholt, S. Mock, I. Altintas, C. Amoreira, Y. Potier, A. Birnbaum, K. Bhatia, and M. Taufer, *Proceedings of the IEEE Special Issue on Grid Computing*, 93 (3): 510-521 MAR 2005

"A Framework for the Design and Reuse of Grid Workflows"
I. Altintas, A. Birnbaum, K.K. Baldridge, W. Sudholdt, M. Miller, C. Amoreira, Y. Potier, B. Ludaescher, *Proceedings of the 1st International Workshop on Scientific Applications on Grid Computing* (SAG 2004), LNCS 3458, 119-132, 2005.

"GEMSTONE: Grid-Enabled Molecular Science through Online Networked Environments. "
Baldridge, K. K.; Bhatia, K.; Greenberg, J.P.; Stearn, B.; Mock, S.; Sudholt, W.; Krishnan, S.; Bowne, A.; Amoreira, C.; Potier, Y.. Invited paper: LSGRID *Proceedings*, 2005.

"Cluster and Grid Infrastructure for Computational Chemistry and Biochemistry"
K. K. Baldridge, W. Sudholt, J.P. Greenberg, C. Amoreira, Y. Potier, I. Altintas, A. Birnbaum, D. Abramson, C. Enticott and S. Garic, *Parallel Computing for Bioinformatics and Computational Biology*, Edited by Albert Y. Zomaya, 2006 John Wiley & Sons, Inc.

Table of Content

List of Tables

List of Figures

1 Introduction and Motivation

1.1 Project Summary

The research project consists of the design and development of quantum chemical based methodology for prediction of structure and properties of molecules in an environment. In this context, consideration of gas phase, solution phase, and crystalline environments are considered.

For consideration of solvent environment, the developed methodology involves a highly accurate implicit solvation model embedded in the quantum chemical modeling program, GAMESS (General Atomic Molecular Electronic Structure Systems). Implementations of key new features of the methodology include a) cavity construction improvement b) a better description of the charge distribution and c) inclusion of non-electrostatic effects. An important goal includes careful consideration of issues associated with first solvation shell effects, that is, non-electrostatic effects involved in the interaction of a solute with the solvent. Through the use of convenient statistical mechanics-based treatments, very accurate solvation properties such as solvation energies, pKa, and polarization charge density can be illustrated. For consideration of crystal environment, the work involves the exploitation of empirical force field to generate possible polymorph crystals in accord with appropriate space groups. Together with an analysis of the intermolecular forces responsible for the packing based on quantum mechanical computations, such studies provide experimentalist with a framework to better understand experimental phenomenon, including exhibited properties.

The described methods are used for prediction in collaborative research involving polycyclic aromatic hydrocarbon systems, especially corannulene based molecules. Corannulene is rich in structural, dynamical and physical features that challenge modern density functional and quantum mechanical methods. Applications of the developed theoretical methods are applied to several topics concerning corannulene-based systems. In these topics, emphasis is placed on structure and dynamical behavior of various functionalized systems in the gas phase, solution phase, and crystal environment.

The thesis is broken down into i) background and motivation, ii) description of pertinent methodology, iii) quantum mechanical-based solvation implementations, iv)

quantum mechanical-based crystal packing theory, v) research investigations, vi) conclusions and perspectives.

1.2 Research Project and Objectives

Considerable efforts have been placed into theoretical chemistry techniques to introduce environmental effects, such as solvent, into the standard computation of structure and properties. Given the importance of solvent, the necessity for an in-depth understanding of its effects in greater detail is clear. Many fundamental questions in biology and chemistry can be better understood when effects of the environment are accurately represented.[1] The goal of the current work is to further development of methodology that can accurately represent the effects of environment, directly in quantum mechanical calculations in the case of solvent, and in a coupled strategy in the case of crystalline environments. The models are discussed in detail, validated, and illustrated through examples associated with important structural, energetic and property effects on key organic systems.

Chapter 2 provides an introduction into theoretical considerations of the details of solvent environments, as well as crystalline environments, both motivated by issues of intermolecular interactions. Chapter 3 introduces the particular QM solvent method addressed in this work, its implementation, and applications in organic chemistry. Chapter 4 provides the description of methodology associated with understanding crystalline environments. Methods are validated and subsequently, in Chapter 5, used for prediction in collaborative research involving polycyclic aromatic hydrocarbon based on corannulene, including properties and dynamics in gas phase, solution phase, and solid state.

2 Computational Methodology and Background

2.1 Theoretical Calculations: Quantum Mechanical Treatment of Environmental Effects

In this work, the term solvation refers to the surrounding of each dissolved molecule or ion by a shell of tightly bound solvent molecules.[2] This solvent shell is the result of intermolecular forces between solute and solvent. The solvation energy is defined as the change in Gibbs energy when an ion or molecule is transferred from a vacuum (gas phase) to a solvent phase. The energy of solvation is the result of contributions from four principal components:

- The cavitation energy: corresponds to the hole that the dissolved molecule creates in the solvent
- The orientation energy: corresponds to the partial orientation of dipolar solvent molecules caused by the presence of solvated molecules
- The isotropic interaction energy: corresponds to long range intermolecular forces (electrostatic, polarization, dispersion)
- The anisotropic interaction energy: corresponds to the energy resulting from formation of hydrogen bonds or electron pair donor/electron pair acceptor interactions, particularly in the first solvation shell.

The dissolution of a substance requires that the interaction energy of the solute molecules (or the lattice energy for a crystal) be overcome, as well as the interaction energy between the solvent molecules themselves. This loss in energy is compensated by a gain in Gibbs energy of solvation, which can be expressed in terms of Gibbs energy of solution and the crystal lattice energy, as shown on the Born-Haber cycle[3] in Figure 1.

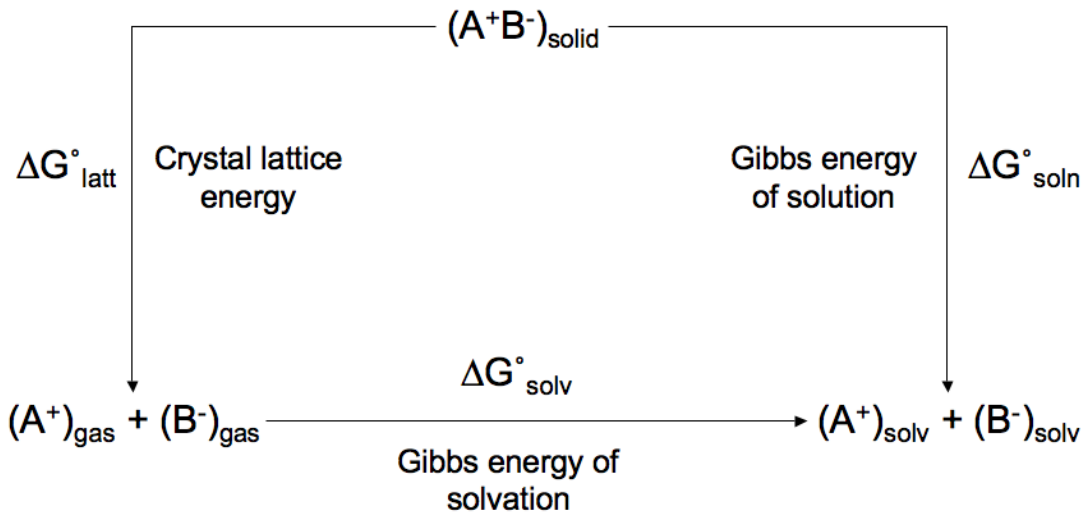


Figure 1: Relationship between standard molar Gibbs energies of solvation and solution, and the crystal lattice energy: ΔGsolv=ΔGsoln - ΔGlatt .

One might ask, to what extent are molecular properties of an isolated molecule altered when embedded in a condensed phase that has its own properties? An aggregate of molecules, in solution or as a solid, is held together by forces derived from the individual molecules that constitute it.[4] Solvation forces are much the same as those constituting solute-solute interactions. The chemistry of recognition describes stabilization and destabilization of molecular aggregates in terms of chemical structure and associated charge distributions. An ideal quantum mechanical calculation would be able to account for all electronic interactions of all species involved, and predict an observed behavior. However, such a calculation is not feasible in practice. The quality of the approximation will define the accuracy of the results obtained. In an attempt to partition the total electronic energies into separate and distinguishable contributions, the estimation is mainly based on molecular structure, polarity and polarizability of the molecules. The remaining intermolecular interactions are typically described partially in classical terms, which are then introduced in quantum mechanical terms.[5]

To ensure a high quality of approximation, an accurate representation of the molecular charge distribution is very important. It has been shown that the coulombic and polarization terms are not enough to fully explain interaction energy between interacting species.[6] Rather, coulombic, polarization-charge transfer, dispersion and repulsion terms are all thought to play an important role, with the representation of these different terms dependent on the actual methodology.

2.2 Quantum Mechanical Treatment of the Solvent

The modeling of structure and reaction processes in solution typically involves one of two strategies - explicit and implicit representation of the solvent environment.

2.2.1 Explicit Solvation Methods

An explicit representation of solvation means treatment of all solvent molecules surrounding a solute of interest directly. Even with current computational capabilities, it quickly becomes computationally expensive to represent the solute as well as all the surrounding solvent molecules in a full *ab initio* manner: the number of degrees of freedom clearly makes the problem computationally prohibitively expensive for most reaction systems of interest. For example, if we treat 200 molecules of water, this adds 1800 degrees of freedom to the computational procedure. Employing quantum chemical methods, the computational effort ranges from $O(N^{2-3})$ to $O(N^{5-7})$, depending on the level of theory considered, making this a limiting determination for feasibility. Typically, when all the solvent molecules are explicitly represented, more appropriate methods are used, such as those based on molecular mechanics force fields (MM, MD, and/or MC). However, in these solutions, one sacrifices considerable accuracy for prediction. Such methods involve a large number of parameters, but also are typically adapted to the description of specific classes of large biochemical system.

2.2.2 Implicit Solvation Methods

A second strategy used to incorporate solvent effects into QM treatments, is to modify the Hamiltonian in the solution of the Schrödinger equation to include the effects of the solution environment. This approach involves the approximation of the solvent as a continuous dielectric surrounding a solute that is contained in a cavity.

Although the creation of a hole (molecule) in the medium (solvent) costs energy (destabilizing effect), direct interactions between the solvent and the solute in the first solvation shell, are stabilizing. The electric charge distribution of the embedded molecule will polarize the medium (induce charge moments), which in turn acts back on the molecule, producing an electrostatic stabilization. Various theoretical strategies vary depending on:

- size and shape of the cavity

- how the charge distribution of the solute is represented
- how the dielectric medium is described
- how the first solvation shell interactions are treated, e.g., nonelectrostatic effects

The solvent medium is represented by a constant value of dielectric, ε, and is the only parameter characterizing the solvent in a continuum method. For this reason, solvents with the same dielectric constant are treated equally: for example acetone (ε=20.7) and 1-propanol (ε=20.1) will be treated more or less the same (different by only .6 dielectric units), even though the hydrogen bonding capabilities are different between them. As such, it becomes very important, in many cases, to consider directed effects of the first solvation shell. In this work, we considered strategies for including these effects in the developed theory presented here.

In 1920, Max Born published work on the free energy of solvation of ions,[7] the solvent surrounding the ion was approximated as a dielectric continuum and a spherical boundary was defined between the ion and the continuum with a corresponding radius depending on the ion. Kirkwood and Onsager subsequently popularized the model for dipolar molecules,[8,9] including the change in dipole from gas phase to solution phase. These methods have been continually improved since that time, and developed theoretical models vary primarily in the definition of the cavity surrounding the solute, treatment of electrostatic effects, and the treatment of nonelectrostatic effects. Different shaped cavities have been used, most simplistically using a spherical shaped, and modified to ellipsoid, and then to more accurate molecular shaped cavities, defined using van der Waals radii for example.[10] The cavity is an important factor as any over- or under- estimation of the area associated with the solvent can drastically change the resulting predicted solvation phenomenon. In general, a molecular shaped cavity is an essential starting point towards accurate agreement with experimental data. However, even with a molecular shaped cavity, a model does not take into account short-range solvation effects: those occurring within the first solvation shell. These contributions, comprised of cavitation energy (entropy and loss of solvent-solvent van der Waals interactions), stabilization due to van der Waals interactions between solute and solvent (also containing a small repulsive component), and repulsion interactions, are very important, particularly if one is trying to attain chemical accuracy.

A alternative application of the continuum model, the Poisson-Boltzmann method[11,12] (Equation 1), employs a second-order differential equation to describe the relation between the electrostatic potential ϕ, the charge distribution ρ and the dielectric constant ε. In the application of the Poisson-Boltzmann, the dielectric constant may depend on the position.

$$\nabla.(\varepsilon(r)\nabla\phi(r)) = -4\pi\rho(r) \qquad \mathbf{1}$$

The numerical aspect of solving the Poisson-Boltzmann equation can make it demanding for use in geometry optimizations of macromolecules.

In quantum mechanical descriptions of the solute M, the calculated electric moment induces charges in the dielectric continuum, which in turn acts back on the molecule, causing the wave function to respond, thereby changing the electric moments. The interaction with the solvent must thus be calculated with an iterative procedure - referred to as a self-consistent reaction field (SCRF) method.[13] For spherical or ellipsoid cavities, the Poisson-Boltzmann equation can be solved analytically but for molecular shaped cavities, it must be done numerically. In these cases, models based on the Born/Onsager/Kirkwood models are less computationally expensive, since a term describing the dipole is added to the Hamiltonian.

As already mentioned, the definition of the cavity in any SCRF approach, strongly influences the results. Unfortunately, there is no real agreement on how to choose the cavity radius. The radii can be derived from experimental data, dipole moment and polarizability. Radii may also be defined based on van der Waals radii of the atoms, scaled with an empirical scale factor to represent the appropriate distance to the solvent molecules. It is also possible to define a cavity based on a contour surface corresponding to the electron density of the wavefunction.

A commonly used QM continuum model, the Polarizable Continuum model[14] (PCM), employs a cavity by interlocking scaled atomic van der Waals radii. PCM uses a detailed description of the electrostatic potential and parameterizes the cavity and dispersion contributions based on the surface area. The basic model used in the present work, based on the original COnductor-like Screening MOdel (COSMO),[15] and modified for QM theory, is another very common model, which has a very detailed cavity construction algorithm, designed to provide very accurate results for

structure and property. The details of this model, as well as the new modifications, will be described in detail in Chapter 3.

Significant advancements in the development of continuum solvation models occurred in the mid-90's, including work in our own group with the COSMO/COSab model to incorporate specific flaws in the general continuum theory.[16] In particular, much work has been done to extend this method to highly accurate methods, such as Coupled-Cluster,[17,18] Møller-Plesset perturbation,[19,20] and density functional theory.[21]

2.2.3 Non-electrostatic Effects

Solvation models include both long-range as well as short-range electrostatic polarizations effects. Effects due to non-electrostatic terms that involve interactions between the solute and the surrounding solvent, albeit a small percentage of the total solvation, can be very important.[22] Historically, the original models of Kirkwood and Onsager considered only the electrostatic effects.

Nonelectrostatic terms are composed of short-range effects, including cavitation, dispersion and the solvent structural effects (CDS); the latter includes both hydrogen bonding and exchange repulsion effects.[23-25] Efforts have been made to integrate nonelectrostatic effects in self consistent reaction field,[10,13] however there is no unique way to separate electrostatic effects from solvents structural effects. Typically nonelectrostatic effects are determined using empirical functions that have been parameterized with experimentally known values, these values are then used to 'correct' free energies of solvation.[26] A common model for the nonelectrostatic contribution to solvation uses the experimental solvation energy of the alkane nearest in size and shape to the molecule of interest. Experimental values for the hydration free energies of alkanes have been compiled in literature.[27]

2.3 Studies in Organic Solids

From decades of research in X-ray crystallography, researchers now have access to a large numbers of crystal structure data for organic compounds. Over the years, there has been an effort from the community to develop methods for the calculation of lattice energies. From this research, force fields have been developed for computer simulation of conformational equilibria in organic molecules.

The total lattice energy of an organic crystal is the sum of the contributions including coulombic, polarization, dispersion and repulsion terms, the strength of which depend on the intermolecular distances. In this section, a brief description of each contribution is provided.

2.3.1 Coulombic Potential Energy

The Coulombic potential energy is stabilizing whenever opposite charges are involved, and destabilizing when charges of the same sign are involved, corresponding to attractive or repulsive forces, respectively.

The Coulombic energy can be expressed in terms of distributed multipole,[28] and is calculated as a sum of moments, monopole, dipole, quadrupole, octapole, etc. Multipolar interaction energies can be stabilizing or destabilizing depending on the respective type and orientation. In solution, molecules have a very fast reorientation and the contact between opposite charge area is more frequent than like-charges, so the total sum is generally stabilizing. However, in a crystal environment, there is no orientation freedom, so even if the Coulombic energies are stabilizing overall, there can be destabilization within the crystal due to packing.[29]

Furthermore, in crystals we need to consider the penetration charge term, corresponding to the correction of Coulombic energy to account for the finite size of a charge distribution. This is necessary for short-range interactions in the crystal, for example to describe typical packing. Otherwise, it can be shown that potential Coulombic energy can be destabilizing at very short range, too short to have any chemical significance.[30] However, as will be shown later, the intermolecular repulsion at short range is independent of Coulombic energies.

2.3.2 Exchange-repulsion

Molecules do not just collapse onto one another. Aside from the long-range repulsion already in play, there are also short-range destabilization effects of a quantum mechanical nature. This destabilization may be explained as the effect of antisymmetrization of the wavefunction. The repulsion arises from the overlap of charge distribution as a consequence of the Pauli principle, which forbids two electrons with the same spin to be at the same point in space simultaneously. When two molecules are far apart, there is no overlap and therefore, spin consideration does

not apply. However, as the electron densities begin to overlap, such effects must be taken into account.

The term, exchange-repulsion, is used due to the relation to the exchange integral, which involves a large negative contribution and is responsible for the primary destabilization in a system, since it contains additional terms depending on the overlap integral.

2.3.3 Dispersion

The small additional stabilization observed in some apolar systems has been described using a coupling of quantum mechanical oscillators. These effects correspond to the electron correlation effects in quantum mechanics, and are related to the oscillating behavior of the electron density. When the density changes, a local dipole is generated that will generate an induced dipole in a nearby electron distribution. The resulting coupling in these dipole-dipole interactions is the basis of the stabilization energetic effect called dispersion.

It is important to note that dispersion effects are a non-classical effect arising from the correlated oscillation of the electrons in the interacting molecules. The dispersion is a consequence of the uncertainty principle, like all zero-point energy effects, and it is purely quantum mechanical effect.

2.3.4 Polarization

Polarization is also referred to as electrostatic induction energy. When an external electric field acts on the electrons of a molecule, the perturbation creates a displacement of the elections under the action of this electric field. This displacement corresponds to the creation of a dipole moment proportional to the electric field, and represents the tendency of a solute electronic distribution to be modified by an external field. Polarizability will be large when electrons are farther from a weakly charged nucleus.

The energy represented by polarization is stabilizing due to induced dipoles aligning in stabilizing directions. However, polarizability is not uniform and if there is a very strong field or extreme polarization, the induced moment is not proportional to the polarization field anymore.

The pioneers first created the atom-atom method.[31] As with most pioneering work, there were problems and room for improvements. The main complication

involved the number of terms obtained when adding all the terms from a lattice summation, requiring several approximations. The method becomes less accurate the higher the number of parameters, however, the flexibility and accuracy increases with an increasing number of parameters. In the end, most of the resulting parameters are determined using experimental data:

- Thermodynamic data: sublimation enthalpy of the crystal is used for comparison with the lattice energy.
- Structural data: in order to obtain a model close to the observed crystal structure, the relaxation of the structure under the action of the potential must be small.
- Vibrational and mechanical data: the potential should be close to the measure lattice vibrational frequencies and the crystal strain and stress tensors.

As such, the success of any approximation for mimicking crystal environments depends on the accuracy of the experimental data. The situation is different from solution phase where a higher mobility brings molecules into collision more often.

Crystal polymorphs are frequently observed as an organic compound crystallizes. As such, it is tempting to use derived crystal theory methods to predict and explain solid-states properties. However, the prediction of possible crystal structures from first-principles, using only the atomic coordinate of the molecular structure, is still controversial.[32] The topic has been addressed many times[33-37] and while promising results have been presented, there is still no definitive treatment available for modeling of polymorphic structure. The problem of crystal prediction is poorly defined because any generated structure can be considered as a potential polymorph. One must keep in mind that, for example, several kinetic issues, equilibrium defect, crystal domains, mosaicity, and internal texture are not explicitly addressed through quantum mechanics techniques. Instead, one is limited to empirical methodologies, where differences in polymorphic forms are measured with respect to their small differences between terms their respective lattice energies.

On the quantum mechanical side, density functional theory (DFT) promises an accurate representation of intermolecular potentials. However, its application to crystals of organic molecules is limited due to system size of appropriate crystalline representations. One promising solution, proposed by Gavezotti, et al., uses

intermediate solution that invokes some QM data, together with an empirical representation of crystalline forces, referred to as the semi-classical density sums (SCDS), or pixel, method.[38] This method can handle intermolecular energies of a crystal for small to medium sized organic molecules (~30 atoms), and requires only a small number of parameters. This theory will be explained and illustrated in detail in Chapter 3.

3 Simulation of Molecular Environment in Quantum Mechanics

3.1 Introduction

Continuum solvation models describe the electrostatic behavior of a solvent by a dielectric continuum.[10] The polarization of the continuum, induced by the charge density of the solute, is represented by the screening charge density appearing on the boundary surface between the continuum and the solvent, i.e. on the surface of the solute cavity. In order to calculate the screening charge density one has to solve the boundary condition problem. Instead of using the exact boundary condition the COSMO "COnductor-like Screening Model" approach[15] applies the simpler boundary condition of the vanishing potential on the surface of a conducting medium. As explain in chapter two, the COSab method is a continuum solvation method based on COSMO. Previous work has shown this method to be highly accurate.[16,39,40]

The general COSab model that is implemented in the GAMESS software has been improved in this work at several levels, including

a. Molecular cavity construction improvements.
b. Treatments of outlying charge: the electrostatic is being treated with different methods to enclose all the electron density: double cavity method (COSab-DC) and distributed multipole (COSab-DM).
c. Strategies to integrate non-electrostatic effects as:
    a. a post-treatment method for COSab (COSab+RS).
    b. direct integration of non-electrostatic effects within quantum mechanics (COSab-D).

3.2 COSab: Fundamental Considerations

The accuracy of Continuum Solvent Method (CSM) primarily resides in the description of the charge distribution and the construction of the cavity (the boundary between the solute and the solvent). The COSab-GAMESS method uses a non-iterative Green's function approach, where the screening charge distribution is expressed as a linear function of the electronic charge distribution. This allows a direct integration into the solute Hamiltonian of the Self Consistent Field (SCF) procedure.

3.2.1 Cavity Construction

The cavity represents the boundary between the solute and the solvent. The algorithm for cavity construction in the COSab-GAMESS method is implemented as follows.[15] The basic idea involves creation of a boundary cavity that follows precisely the shape of the molecule in question. This method uses a radius RSOLV, which is atomic van der Waals radii increased by 20%, in accord with common findings in the literature,[41,42] to generate this surface.

The construction of the cavity begins with a basic grid is generated from an iterative refinement of triangles, beginning from a regular icosahedron (Figure 2). This is followed by a series of projections onto the solvent surface, extending the cavity to approximately a distance of van der Waals + 20%.[41,42] Small enough segments are created on the surface as to ensure the assumption of homogeneous charge distribution on that segment. Grid points are assigned to the closest segment, and grid points that are inside neighboring spheres are excluded. The value defining the number of segments per atom is a parameter since the charge distribution depends on the system of interest; specifically, on the electronic distribution and particularly molecular structure.

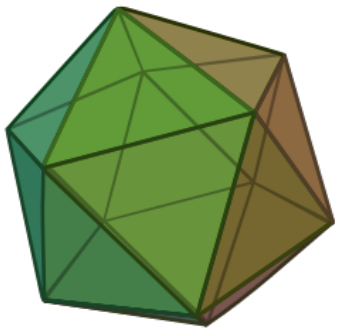


Figure 2: Regular icosahedrons.

The resulting cavity is represented with m surface segments, the centers of which makeup a grid of homogeneous points. Each segment has a corresponding potential, $\Phi_1 ... \Phi_m$, related directly to the charge distribution of the solute. The screening charges, q_m , which arise from the polarization of the continuum, are calculated from the directly integrated QM potential, Φ. These charges screen perfectly the charge of the solute, Q. The potential used to calculate these charges depends on the methods for correction of outlying charge error that has been chosen. The coulomb interaction matrix, A_{ij}, of these screening charges, takes into account the

self-interaction of a homogenous charge distribution on a segment by the diagonal element A_{ii}.

Cavity construction is the critical component of accuracy of any continuum solvation method. As such, the cavity construction methodology is one distinguishing feature of the different continuum methods (CSMs), and thus their performance.[10] The cavity construction algorithm in the COSab-GAMESS implementation has been improved in this work to increase overall accuracy of the method.

The quality of the cavity depends significantly on segmentation granularity. Since the cavity surface is describe as homogenous segments: a single charge can represent a segment. The value of discretization, representing the number of segments per atom, is therefore important because if the segment is too large the assumption of homogeneity breaks down. The density of the default grid is 1082 points per unit sphere. The default value for the segment construction (discretization value) is 92 segments per unit sphere. These defaults settings have been shown to be sufficient for good accuracy,[16] but may be increased as needed for different systems.

3.2.2 Solute/Solvent Description

COSMO follows the concept of conductor, where one considers the ideal screening charges, q*, that would arise in a conducting continuum environment. With a perfect conductor of infinite permittivity (ε=∞), the total potential arising on the surface segments due to the solute and the screening charges has to vanish. This assumption greatly simplifies the mathematics of the problem such that the potential caused by screening charges, q, is given by Aq:[16]

$$0 = \Phi + Aq^* \Rightarrow q^* = -A^{-1}\Phi \qquad \mathbf{2}$$

However, solvents do not behave like perfect conductors, so the screening charges are approximated using the dielectric constant of the medium with:

$$q = f(\varepsilon)q^* \qquad \mathbf{3}$$

With

$$f(\varepsilon) = \frac{(\varepsilon - 1)}{(\varepsilon + 0.5)} \qquad \mathbf{4}$$

After calculating these charges on the cavity surface from knowledge of the gas phase wavefunction, they are added back into the 1-electron Hamiltonian in the quantum mechanical procedure. The interaction energy between the solute and the screening charges is

$$E_{INT} = \Phi q \qquad \mathbf{5}$$

where Φ is the electrostatic potential arising on the surface from the charge distribution of the solute. Half of this energy is used for the creation of the cavity, so the total energy gain of the system is define as follow:

$$E_{DIEL} = \frac{1}{2}\Phi q = -\frac{1}{2} f(\varepsilon)\Phi A^{-1}\Phi \qquad \mathbf{6}$$

This is followed by continuing the self-consistent iterative procedure using the solute density in the presence of the screening charges. The energy of the entire system is defined in Equation 7, where E_0 is the SCF energy of the solute alone.

$$E_{SCF} = E_0 + \frac{1}{2}\Phi q \qquad \mathbf{7}$$

3.2.3 Outlying Charge Error Treatment

In general, the cavity construction is not based on the wavefunction of the solute, but on the coordinates of the molecule. As such, there is a possibility that some of the electron density is located outside the constructed cavity, particularly if the molecular system of interest is very diffuse in charge. This density, denoted as "outlying charge error", can represent a significant amount of energy, particularly for diffuse molecular systems, such as anions. There are currently two methods for accounting for outlying charge, as implemented in the COSab method, and both have been shown to provide an improvement in the prediction of solvation phenomenon.[40]

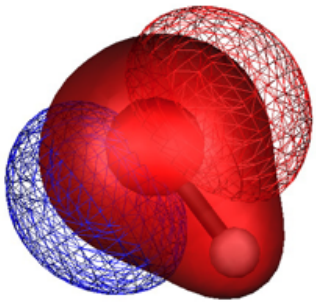


Figure 3: Exaggerated view illustrating wavefunction outside the cavity built around the solute: example of the hydroxide anion.

The theory and general equations involved in the distributed multipole (DMULTI) method as well as the double cavity (DBLCAV) method will be outlined in the following sections and only briefly described here. The distributed multipole method (DMULTI) relates the directly integrated solvent-influenced potential of the

solute, to a representation of the potential in terms of atom-centered distributed multipoles, the latter of which is used to evaluate the revised charges in the iterative procedure since it includes the effects of the solvent on the entire electronic structure.[16] The difference in the two representations of potential is considered the OCE. The double cavity method simply defines a secondary cavity constructed beyond the original molecular shaped cavity, which is presumed to fully include any of the potential beyond the original cavity wall. The outlying charged is then assessed from the difference of potential between the two cavities.[40]

The literature presents both approaches of OCE, and provides illustration of their accurate predictions of solvent effects.[16,20,40,43] This work compares the two methods in light of the work carried out in this thesis, in particular with reference to[43]:

a) enhancements made to the cavity generation, and

b) new strategies for prediction of solvation phenomenon, including first solvation shell effects.

3.2.3.1 Distributed Multipole Method: COSab-DM

The distributed multipole method[44] use a representation of Q as a set of k atom-centered multipoles, M(Q), up to hexadecapoles in order to represent the solute charge density.[28,44] The potential, Φ, is therefore represented as BM(Q) arising on the m segments from the k multipoles, B being the matrix (k x m) corresponding to the Coulomb interaction of the multipoles with the segments. The resulting pseudo-potential Φ_{DM} is used to calculate the screening charges that will go into the SCF procedure.

$$\Phi^{DM} = BM(Q) \qquad \mathbf{8}$$

The screening charges are calculated from this potential:

$$q^{DM} = -A^{-1}\Phi^{DM} \qquad \mathbf{9}$$

The energy resulting from the SCF loop is therefore:

$$E_{SCF} = E_0 + \frac{1}{2}\Phi q^{DM} \qquad \mathbf{10}$$

The screening charges are calculated from the multipole potential that is insensitive to OCE. The SCF procedure also allows taking into account the back-polarization of the solute by the solvent. Since the charge density is represented inside the cavity, we eliminate almost all outlying charge. But we are using Φ, the real

solute potential to calculate the energy, there is still some error associated with outlying charge in Equation 11 and it needs to be corrected:

$$OCE^{DM} = (\Phi q^{DM} - \Phi^{DM} q^{DM}) \quad \textbf{11}$$

The total energy needs to be corrected according to Equation 12:

$$E_{TOT} = E_{SCF} - \frac{1}{2} OCE = E_0 + \frac{1}{2}\Phi q^{DM} - \frac{1}{2}(\Phi q^{DM} - \Phi^{DM} q^{DM})$$

$$E_{TOT} = E_0 + \frac{1}{2}\Phi^{DM} q^{DM} \quad \textbf{12}$$

3.2.3.2 Double Cavity Method: COSab-DC

In the method of the double cavity, a second molecular shaped cavity is defined outside the original cavity in order to enclose most of the electron density of the solute.[40] The points on outer cavity have a one to one correspondence with the points on the inner cavity. The outer cavity segments charges q' are calculated and then added to those of the inner cavity. The treatment of the OCE is only post-SCF in this method and the SCF energy is calculated from Equation 7. The potential of the outer cavity and the screening charge of the inner cavity is calculated, Φ_{DC}. From there we calculate the charges q_{DC} according to:

$$q^{DC} = -(A^{DC})^{-1}\Phi^{DC} \quad \textbf{13}$$

where A^{DC} is the interaction matrix of the outer cavity points. The corrected screening charges, noted q", are calculated by adding the charges of the outer cavity to those on the inner cavity. This is possible because of the one to one correspondence of the points of both cavities.

$$q'' = q + q^{DC} \quad \textbf{14}$$

After calculating the outlying charge, we can correct the potential:

$$\Phi'' = -q'' A \quad \textbf{15}$$

where A is the interaction matrix between the point of the inner cavity. We can correct the same way the dielectric energy:

$$E_{DIEL} = \frac{1}{2}\Phi'' q'' \qquad \textbf{16}$$

The outlying charge error with the double cavity method is written as:

$$OCE^{DC} = (\Phi q - \Phi'' q'') \qquad \textbf{17}$$

The total energy of the system is therefore corrected as:

$$E_{TOT} = E_{SCF} - \frac{1}{2} OCE^{DC} = E_0 + \frac{1}{2}\Phi q - \frac{1}{2}(\Phi q - \Phi'' q'')$$

$$E_{TOT} = E_0 + \frac{1}{2}\Phi'' q'' \qquad \textbf{18}$$

3.2.4 Implementation inside GAMESS

Both of the described outlying charge error methods are implemented inside the GAMESS package.[45] The COSab method is available with most of the ground state conventional *ab initio* methods: restricted and unrestricted Hartree Fock (RHF, ROHF, UHF)[46,47] and MP2 (Møller-Plesset perturbation theory).[48,49] COSab can also be used with density functional theory based on Kohn-Sham orbitals.[50-55] Single point (energy) calculations are available as well as gradient and second derivative calculations, enabling geometry optimizations, Hessians, and reaction paths, within the COSab algorithm for all of these available methods.

A schematic of the COSab implementation inside the GAMESS software is illustrated in Figure 4.

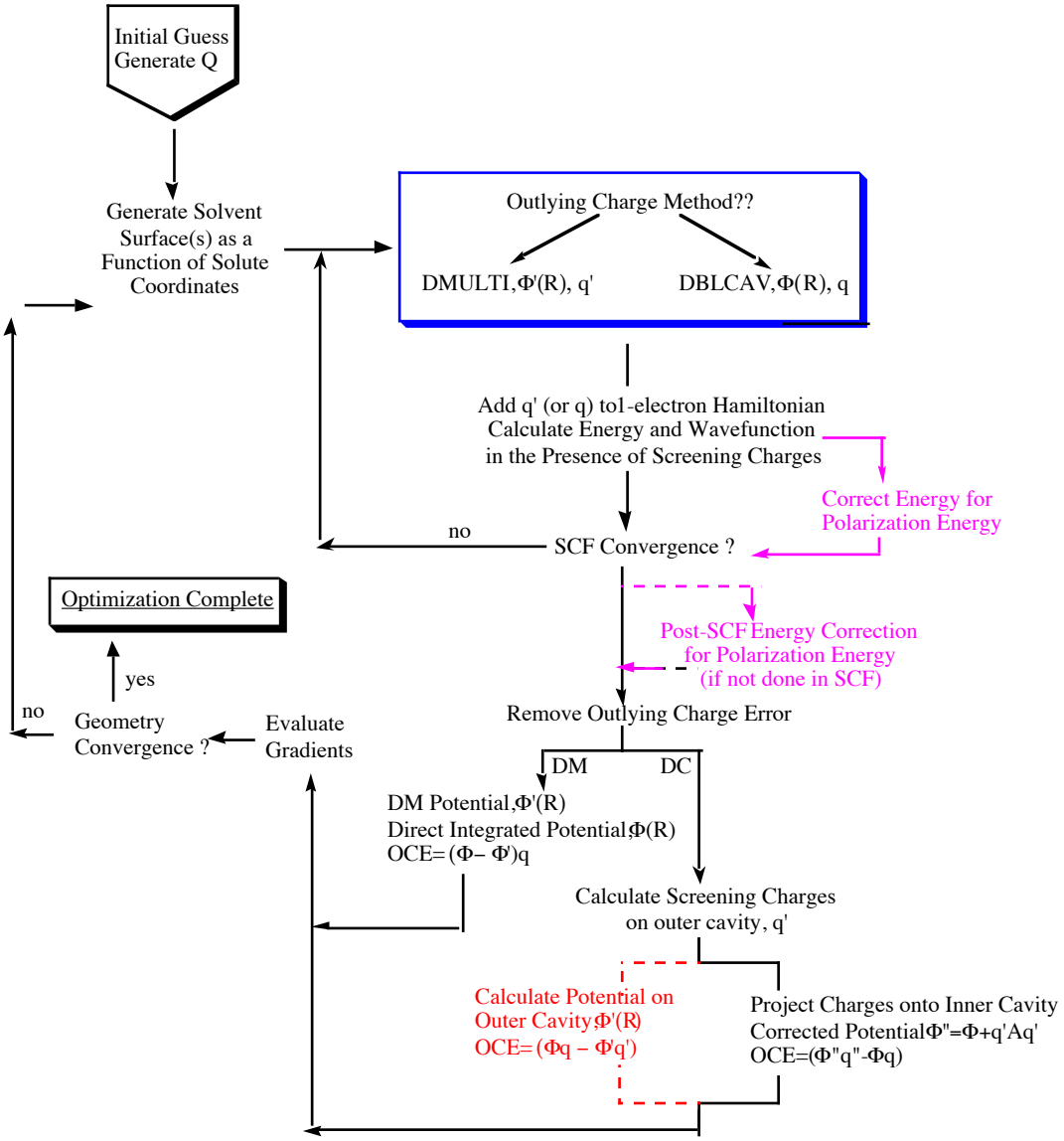


Figure 4: Flow diagram for COSab method within the Self-Consistent procedure in GAMESS.

The current work involved the following updates to this overall algorithm:

a) Enhanced cavity construction.

b) Enhanced and more efficient description of the electrostatics (including the OCE).

c) Strategies are implemented for inclusion of non-electrostatic effects.

3.3 Enhanced Cavity Construction & Resulting Electrostatic Solvation Predictions

3.3.1 Cavity Closure

A notorious difficulty for many continuum solvent models is the handling of molecules where the distance between any two atoms is close to the sum of their individual radii, because the construction of the solvent surface is limited by the contact of interacting spheres. This situation can happen in cases of molecules having irregular shaped, highly varying surface contour, or in a reaction process when bonds and angles are stretched, as in transition states. In these cases, simple cavity constructions can result in misrepresented molecule-solvent interactions in the model. When the distance between two atoms is close to the sum of their radii, the atoms spheres just slightly intersect, creating a deep crevice in the surface. The cavity construction algorithm has been enhanced to consider these hidden parts of the surface. A previously described algorithm of 'cavity closure' [43] has been implemented in the current version of GAMESS, that accounts for the potential loss of accuracy. The original crevices are closed by sets of interconnecting triangles. The total number of segments may be increased by 50% with this inclusion, providing a more homogenous-segmented surface. The computational cost associated with this enhancement is minimal. In contrast, there are known methods that solve these same problematic surface contours by constructing additional spheres to mask the crevices, which then significantly increase the computational cost, (e.g., PCM/GEPOL cavity construction).[56]

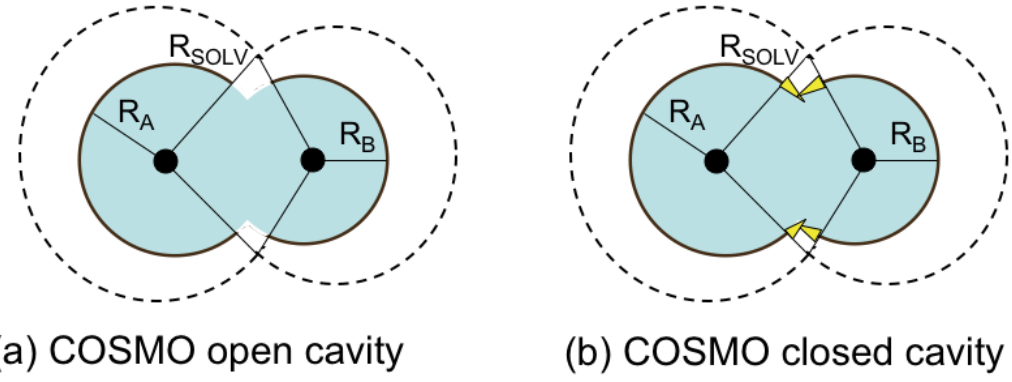


Figure 5: Schematic illustration of the cavity construction.[57]

Using the enhanced method for cavity closure, a comparison between the free energy of solvation with and without closure on a set of neutral organic molecules has been done and the results reported in Table 1. The column labeled ‘closure’ use the closed cavity algorithm and ‘non-closed’ use the open cavity. The computed surface area and volume of the resulting cavity are reported in Table 2. The data from these tables are represented in Figure 6 and Figure 7 respectively.

Table 1: Electrostatic energy of solvation, ΔE_{elect} (kcal/mol), with closure and without cavity closure, for two methods of outlying charge error correction, as compared to experiment.

Molecule	ΔE_{elect} , Double Cavity, DC, outlying charge error (OCE) method					ΔE_{elect} , Distributed Multiple, DM, outlying charge error (OCE) method					
	Closed Cavity			Unclosed Cavity		Closed Cavity			Unclosed Cavity		
	ΔE_{elect}	ΔE_{elect} (expt-calc)	OCE	ΔE_{elect}	ΔE_{elect} (expt-calc)	ΔE_{elect}	ΔE_{elect} (expt-calc)	OCE	ΔE_{elect}	ΔE_{elect} (expt-calc)	ΔG expt
acetic acid	-7.86		-0.21	-7.51		-7.75		0.06	-7.20		
acetone	-6.07	2.17	-0.16	-5.91	2.01	-6.00	2.10	0.21	-5.47	1.57	-3.90
carbon tetrachloride	-1.52	0.42	-0.05	-1.36	0.26	-1.66	0.56	-0.14	-1.43	0.33	-1.10
methane	-0.37	2.37	0.03	-0.36	2.36	-0.32	2.32	0.06	-0.29	2.29	2.00
ethane	-0.37	2.17	0.03	-0.36	2.16	-0.32	2.12	0.04	-0.27	2.07	1.80
formaldehyde	-5.28	.72	0.01	-5.20	.80	-5.23	.77	0.11	-5.07	.93	-7.00
water	-7.25	0.95	-0.13	-7.18	0.88	-7.21	0.91	-0.04	-7.10	0.80	-6.30
hydrogen fluoride	-5.54	.06	-0.08	-5.44	.16	-5.47	.13	0.3	-5.33	.27	-5.60
methanol	-5.20	0.10	-0.13	-5.04	.06	-5.14	0.04	0.04	-4.83	.27	-5.10
ammonia	-5.28	1.08	-0.07	-5.24	1.04	-5.26	1.06	-0.08	-5.23	1.03	-4.20
propadiene	-4.99		-0.25	-4.77		-4.94		-0.04	-4.51		
propane	-0.45	2.45	0.04	-0.43	2.43	-0.40	2.40	0.02	-0.33	2.33	2.00
triethylamine	-2.30	.90	-0.09	-2.15	.05	-2.27	.93	0.00	-1.94	.26	-3.20

Table 2: Surface (A^2) and Volume (A^3) with, C, and without, NC, cavity closure.

	Surface		Volume	
Molecules	NC	C	NC	C
acetic acid	224.9	326.2	1066.5	495.8
acetone	239.5	363.1	1188.1	576.6
carbon tetrachloride	356.3	472.8	2001.6	853.0
methane	159.3	199.0	619.4	254.6
ethane	189.7	273.3	847.4	392.1
formaldehyde	178.9	218.2	725.5	289.0
water	128.7	151.9	438.6	170.1
hydrogen fluoride	130.4	141.2	441.4	155.5
methanol	176.0	238.8	735.6	322.4
ammonia	139.5	173.1	502.2	205.0
propadiene	230.2	314.9	1071.4	467.8
propane	219.4	344.1	1074.1	536.6
triethylamine	241.8	391.3	1254.8	649.7

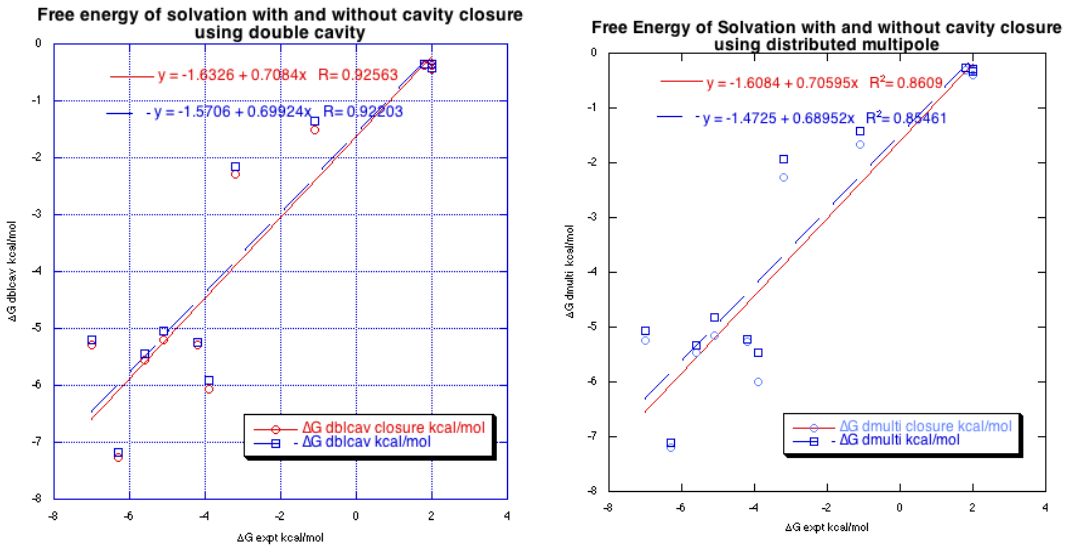


Figure 6: Free energy of solvation comparison between cavity closure methods for a set of neutral molecules.

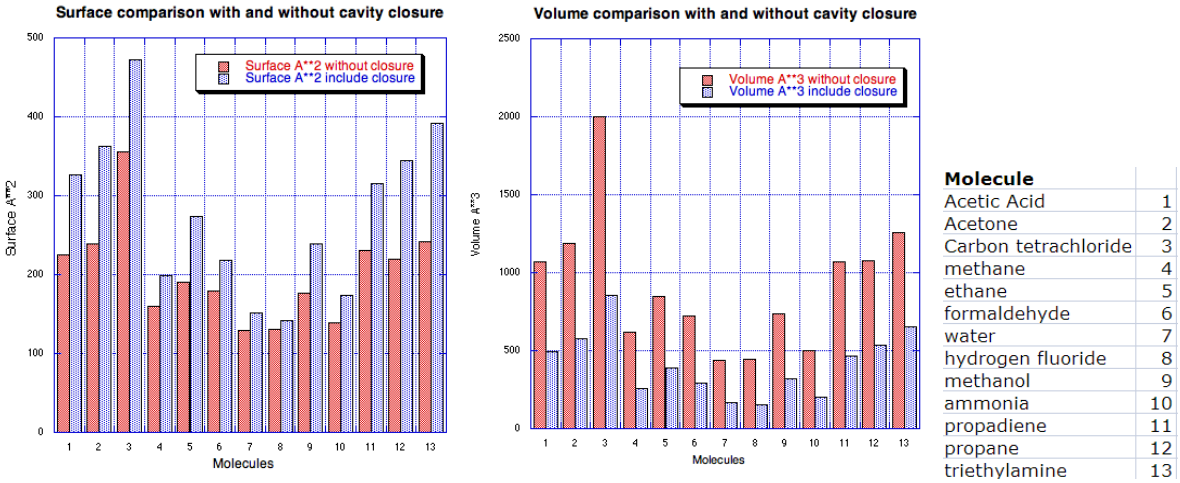


Figure 7: Surface (A^2) and volume (A^3) differences for a set of neutral molecule with and without cavity closure.

The enhanced cavity construction can be very important for realistic solvation modeling in general.[43] On the set of small molecule, we notice a slight improvement on the quality of the free energy of solvation using the enhanced cavity. The improvements in the evaluation of the energy of solvation using the cavity closure method are small on this particular set of molecules, representing only a few kcal/mol of difference at the most. Given the relatively high accuracy of the method even without the closure procedure, this might be anticipated, particularly for this set of small molecules. This improved cavity description, however, is expected to be greater for molecule with more problematic shapes or charge distributions, than what is considered in this testset.[43]

Given this, one can still analyze the behavior of the different methodologies with the cavity correction. To the extent that the correction will represent a larger exposure of the surface to the solvent environment, the predicted values will tend to be larger than values predicted without the correction, as can be seen by inspection of the ΔE_{elect}(expt-calc) values in Table 1. Comparing only the differences in OCE methodology, one finds that the cavity correction appears to be more significant in the DMULTI method. This is likely a consequence of the OCE procedure, since in the DBLCAV method, the OCE method involves the use of a second cavity to estimate and eliminate any wavefunction extent outside the molecular cavity. In this way, effects of closure and effects of OCE are somewhat mixed, potentially resulting in some cancellation of error. In the DMULTI method, the wavefunction extent is eliminated via a distributed multipolar representation, and the estimated contracted

wavefunction will naturally provide a larger interaction energy in the closure points. The most dramatic difference in the set of molecules is observed with acetone, where for the DMULTI method, one finds a difference of over ½ kcal/mol between the closed and non-closed cavity results.

The volume and surface changes (Figure 7) provide another indication of the extent of the surface modification with and without the closure. With inclusion of the closure, the surface area increases, due to the increase in number of surface segments used for description of the boundary. However, the total volume actually decreases, since by putting in the extra interaction areas, you are excluding volume that should not be described as part of the molecule, but as part of the solvent. Depending on the actual molecule, one can see that this extra volume can be quite extensive, and in cases where there is highly varying charge in these regions, one would expect a significant variation in the predicted solvent phenomenon. For example, in the case of tetrahedral molecule, such as carbon tetrachloride or methane, the volume of the cavity is smaller by a factor of two (2001 vs. 853) using the new cavity construction, due to the tight regions around the carbon in the tetrahedral shape. The resulting increase in number of surface segments needed to accommodate this description then results in an increase in the surface area value from 356 to 472. Alternatively, a molecule with essentially no crevices, such as HF, shows very little increase in surface area.

Instead of creating more spheres, as explain earlier, the updated algorithm involves filling these areas of sharp contours and crevices with an additional set of surface segments that can still fulfill the constraint of homogeneous charge on that surface segment. In all cases, having the right description of the cavity around the molecule is a can be an important factor for prediction of chemical behavior.

3.3.2 Discretization Effect

As previously mentioned, another underlying assumption in the cavity construction is that the segments making up the cavity are assumed to be homogeneous in charge. Since the cavity is based on the atomic coordinates of the solute atoms, when a geometry optimization is carried out, for each new geometry the molecular cavity must be re-constructed and the associated charges re-calculated. One must find the right balance between the extra cost of the cavity creation and the

necessary refinement of the segmentation to ensure homogeneity in charge on each segment.

The value of discretization represents the number of segments per atom, and therefore controls the size of the segments on the cavity surface. If the segment size is too large, the assumption that the charge distribution is homogeneous on the segment breaks down. The value NSPA represents the discretization, or the number of segments per atom. This parameter controls the size of the segments on the cavity surface, in that the larger the value for NSPA, the smaller the individual segment surface area is. The basic points are constructed to achieve the most compact segments.

Table 3 shows the variation of electrostatic solvation energy as a function of the size of the cavity discretization for 3 different molecules, for both DMULTI and DBLCAV OCE implementations.

Table 3: Effects of discretization on RHF/6-31G(2d,p)//RHF/6-31G(2d,p) solvation energies for CH4, H2O and HF. Values in parenthesis are experimental free energies of solvation (kcal/mol).

Discretization	CH_4 (2.00)		HF (-5.60)		H_2O (-6.30)	
	ΔE_{elect} DBLCAV	ΔE_{elect} DMULTI	ΔE_{elect} DBLCAV	ΔE_{elect} DMULTI	ΔE_{elect} DBLCAV	ΔE_{elect} DMULTI
12	-0.49	-0.38	-6.22	-6.02	-8.83	-8.52
32	-0.47	-0.38	-6.03	-5.93	-8.28	-8.26
42	-0.46	-0.38	-6.01	-5.92	-8.22	-8.22
62	-0.46	-0.38	-6.01	-5.92	-8.22	-8.22
92	-0.41	-0.35	-5.74	-5.71	-7.82	-7.94
132	-0.40	-0.35	-5.73	-5.69	-7.76	-7.89
162	-0.40	-0.35	-5.72	-5.69	-7.74	-7.87
362	-0.38	-0.34	-5.62	-5.62	-7.61	-7.78
482	-0.38	-0.34	-5.61	-5.61	-7.59	-7.76

We can see from Table 2 that, for these simple molecules, a moderate number of segments (e.g., NSPA=92) provides optimal accuracy. On the other hand, very low values of discretization (e.g., NSPA=12), can result in significant error in predicted solvation energies. For example, the experimental value of free energy of solvation of hydrogen fluoride is -5.60 kcal/mol; when the number of segments is minimum the calculated energy of solvation is -6.22 kcal/mol, compared to even the average value of 92, where the solvation energy is less than 0.15 kcal/mol than the experimental value, and less than 0.05 kcal/mol with the highest value of discretization. In the case of water, one finds a converged value at the highest discretization value that is still

greater than 1 kcal/mol from the experimental value, as also is the case for methane. This potentially points to the limits of the method without explicit inclusion of first solvation shell effects, which are not included as of yet.

One can additionally see the importance of discretization value on accuracy as a function of polarity of the system. While a non-polar system, such as CH_4, does not appear to significantly vary in predicted solvation energy as a function of discretization energy, for a highly polarized system, such as HF, the effect of discretization is shown to be very important (e.g., here shown to amount to ~1 kcal/mol. Previous examples from this group has shown that, for more complex structures, like hydroxymethyl acylfulvene, the effect can be more dramatic, and require a reasonably high value of NSPA to provide consistent results.[43] It is also often the case that the calculation will not have good converge properties for the optimization procedure with low values of discretization. On the other hand, as can be seen in the data in Table 3, beyond a certain sufficiently high value of discretization, one see complete convergence, and use of very high values only results in unnecessary computational expense. We generally recommend an optimal value of discretization of 92 segments per sphere for medium size basis sets, and somewhat larger for more sensitive basis sets particularly those including diffuse functionality.

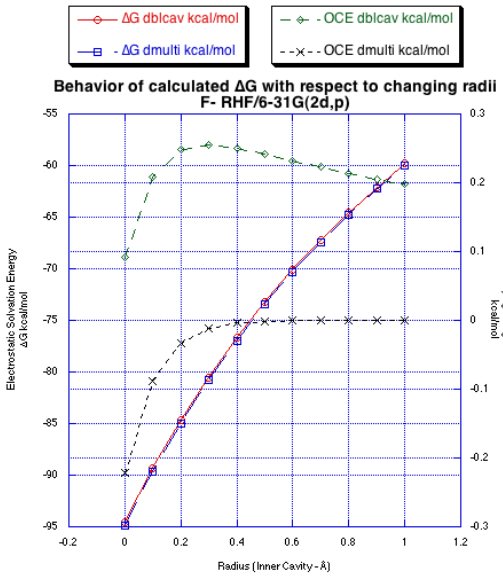
3.3.3 General Radial Dependence

Since the radii of each atom is used for defining the cavity, and the position of the charges depends on this cavity extent into the solvent, one can imagine that there is a significant radial dependence on the prediction of ΔG_{elect} with respect to the radius of the cavity.[57] For example, an exact r^{-1} behavior should be obtained from radially symmetric ions (monopolar). A r^{-3} behavior should be observed for dipolar neutral molecules, a R^{-5} for quadrupolar molecules, and R^{-7} behavior for octopolar molecules, etc. This behavior can be affected by the exponential decay contribution of the outlying charge effect, if they are not correctly accounted for. The effects of radii on results of ΔG and OCE is illustrated for the current model on two ions, K^+ and F^-, and one dipolar neutral molecule, H_2O, and summarized in Table 4.

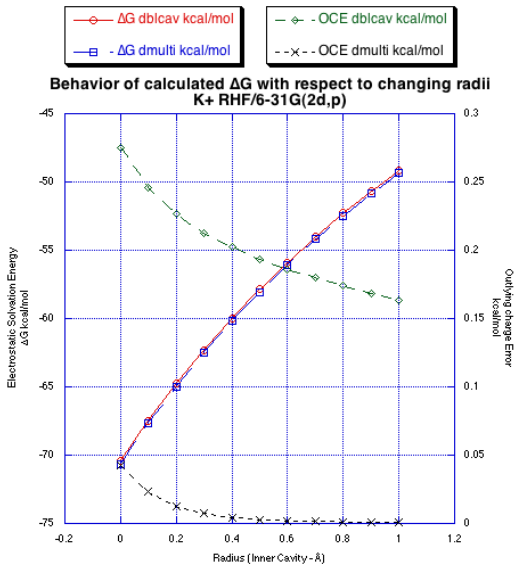
Table 4: RHF/6-31G(2d,p) solvation energies and outlying charge correction as a function of surface radii for selected systems. The initial radii used are the optimized radii of Klamt and Jonas[42] systematically increased by 0.1 Å, to illustrate the radial behavior.

	K+				F-				H2O			
Radius	ΔE_{elect} dc	OCE dc	ΔE_{elect} dm	OCE dm	ΔE_{elect} dc	OCE dc	ΔE_{elect} dm	OCE dm	ΔE_{elect} dc	OCE dc	ΔE_{elect} dm	OCE dm
0	-70.39	0.2748	-70.62	0.0425	-94.54	0.0915	-94.85	-0.2219	-7.25	-0.0506	-7.21	-0.0268
0.1	-67.47	0.2460	-67.69	0.0234	-89.34	0.2084	-89.64	-0.0878	-6.56	0.0145	-6.48	0.0813
0.2	-64.78	0.2267	-65.00	0.0129	-84.69	0.2478	-84.97	-0.0331	-5.97	0.0297	-5.89	0.0973
0.3	-62.30	0.2128	-62.51	0.0072	-80.50	0.2551	-80.76	-0.0119	-5.35	0.0424	-5.29	0.0958
0.4	-60.00	0.2022	-60.20	0.0041	-76.70	0.2503	-76.96	-0.0041	-4.75	0.0281	-4.69	0.0862
0.5	-57.87	0.1936	-58.06	0.0025	-73.25	0.2416	-73.49	-0.0013	-4.15	0.0354	-4.13	0.0599
0.6	-55.88	0.1863	-56.06	0.0017	-70.09	0.2321	-70.32	-0.0004	-3.50	0.0293	-3.49	0.0393
0.7	-54.02	0.1798	-54.20	0.0013	-67.19	0.2228	-67.42	-0.0001	-2.95	0.0199	-2.95	0.0221
0.8	-52.29	0.1738	-52.46	0.0011	-64.53	0.2140	-64.74	0.0000	-2.52	0.0142	-2.52	0.0124
0.9	-50.66	0.1684	-50.82	0.0010	-62.06	0.2059	-62.27	0.0000	-2.17	0.0105	-2.18	0.0069
1	-49.13	0.1633	-49.29	0.0009	-59.78	0.1983	-59.98	0.0000	-1.89	0.0082	-1.89	0.0037

a)



b)



c)

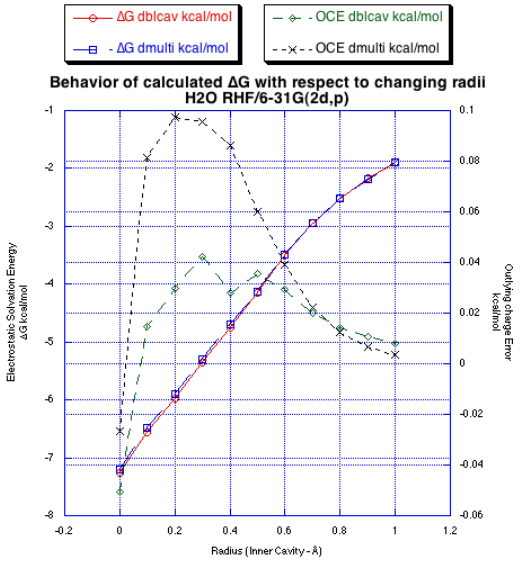


Figure 8: RHF/6-31G(2d,p) calculated ΔE_{solv} and OCE radial dependence for a) F^-, b) K^+, and c) H_2O. Both strategies of outlying charge are considered, DMULTI and DBLCAV methods.

The behavior of the solvation energy reveals the expected radial dependence, as illustrated in Figure 8a and Figure 8b, with r^{-1}, for symmetric (monopolar) ions, F^- and K^+, and in Figure 8c with r^{-3} behavior for the dipolar neutral molecule. H_2O. One can also see the difference in variation in outlying charge effect as a function of radii. In particular, one notices a clear difference in the two solvation implementations and corresponding treatments of outlying charge for simple ions. In the distributed

multipole technique, DMULTI, there is a much smaller error associated with outlying charge at the onset, and a rapid decrease in value being essentially eliminated at 0.4 Å. On the other hand, the DBLCAV approach has a much large outlying charge error associated with it, and a much slower degradation with radius. In contrast, the two methods are much more comparable for treatment of the water molecule, and appear to be the more regular trend, as will be seen in the data below. In either method, however, the full estimate of OCE within the model is eliminated in the algorithm, and so the resulting predicted ΔE_{elect} is effectively the same in either case.

Table 5: Optimized solvent radii used with COSab.

Atom	COSab radius	Bondi radius
H	1.3	1.2
C	2	1.7
N	1.83	1.55
O	1.72	1.52
F	1.72	1.47
Cl	2.05	1.75
Br	2.16	
I	2.32	
S	2.16	1.8

The solvent radii used in the COSab algorithm are presented in Table 5. These radii have been optimized by Klamt and Jonas,[42] with respect to experimental solvation energies of a large class of known compounds. Except for hydrogen, these radii are 13-18% larger than the corresponding van der Waals radii, and thus agree with the widely accepted view of "van der Waals plus 20%" radii within continuum model approaches. Note the difference in these values, for example, from that of the well-known Bondi radii. The parameterized solvent values are more appropriate to the theory of solvation, since the molecular interactions with the surrounding solvent happen within the extended van der Waals region.

3.4 Model Validation

3.4.1 General GAMESS-COSab Methodology

All calculations have been carried out using the GAMESS software package[45] with a variety of wavefunction methods and a wide range of basis set. The wave-function-based methods include Hartree-Fock (HF), second-order Møller-Plesset perturbation theory, density functional theory (DFT), and hybrid DFT. Basis sets used include the following: STO-3G, 3-21G, 6-31G(nd,mp), 6-311G(2d,p),DZV(d),TZV(nd,mp) and cc-pVDZ.[48-50,58,59] The solvation methods show both DMULTI and DBLCAV strategies, and the cavity employing the new closure algorithm. A dielectric permittivity value of ε=78.4 is used for water for all calculations except if otherwise stated. The generation of the cavity construction used 1082 points for the basic grid, 92 segments on a complete sphere, and a solvent radius of 1.3Å. Atomic radii, as optimized in previous work, are employed.[41]

3.4.2 General Behavior of COSab: Neutral and Ionic Molecules

Calculated COSab results for 13 neutral molecules, 7 cations, and 8 anions, with varying structural functionalities are shown in Table 6 to Table 9. Calculated ΔE_{elect}, OCE are all presented. Experimental ΔG values for all available examples are shown for references. Figure 10 to Figure 13 also shows the calculated solvation energies, using both outlying charges methods (from data in Table 6 to Table 9), plotted against the experimental free energy of solvation.

Table 6: RHF/6-31G(2d,p) electrostatic solvation energies and outlying charge vs. experiment. ΔE_{elect} dc is calculated using double cavity method, ΔE_{elect} dm is calculated using distributed multipoles method. G(expt-calc) is the difference between experimental and calculated free energy (kcal/mol).

Molecules	ΔE_{elect} dc	ΔE_{elect} dm	ΔG expt	OCE dc	OCE dm	G(expt-calc) dc	G(expt-calc) dm
CH_3COOH	-7.86	-7.75		0.0175	0.0547		
$CH_3C(O)CH_3$	-6.07	-6.00	-3.9	0.2119	0.1352	-2.16	-2.10
CCl_4	-1.52	-1.66	-1.1	-0.0678	-0.2682	-0.41	-0.56
CH_4	-0.37	-0.32	2	0.0252	-0.0218	-2.37	-2.31
C_2H_6	-0.37	-0.32	1.8	0.0332	-0.0656	-2.17	-2.11
$HC(O)H$	-5.28	-5.23	-7	0.1409	0.1442	1.72	1.76
H_2O	-7.25	-7.21	-6.3	-0.0532	-0.0272	-0.95	-0.90
HF	-5.54	-5.47	-5.6	0.0037	0.0719	0.05	0.12
CH_3OH	-5.20	-5.14	-5.1	0.0395	0.0212	-0.10	-0.04
NH_3	-5.28	-5.26	-4.2	-0.0479	-0.0814	-1.07	-1.06
CH_2CCH_2	-4.99	-4.94		-0.0525	-0.1089		
$CH_3CH_2CH_3$	-0.45	-0.40	2	0.0404	-0.1223	-2.44	-2.40
$(CH_3)_3N$	-2.30	-2.27	-3.2	0.0666	-0.1236	0.90	0.93
CH_3^+	-78.85	-78.83		0.9379	0.9312		
NH_4^+	-81.88	-81.72	-77	0.8571	0.9866	-4.87	-4.72
$CH_3NH_3^+$	-72.63	-72.53	-68	1.1526	1.1871	-4.63	-4.53
$(CH_3)_2NH_2^+$	-65.17	-65.12	-61	1.4090	1.3500	-4.17	-4.12
$(CH_3)_3NH^+$	-58.75	-58.83	-57	1.6742	1.4520	-1.75	-1.83
$(CH_3)_3O^+$	-56.70	-56.69		1.6036	1.4958		
K^+	-70.39	-70.62		0.2747	0.0424		
OH^-	-97.69	-97.82	-106	-1.3446	-1.5058	8.30	8.17
CN^-	-75.25	-75.42	-77	-1.3606	-1.5553	1.75	1.57
NO_2^-	-73.74	-73.95	-72	-0.9066	-1.1213	-1.74	-1.94
NH_2^-	-91.69	-91.83	-93	-2.543	-2.7836	1.31	1.16
BH_4^-	-74.72	-75.13		-4.7453	-5.5577		
BF_4-	-62.29	-62.45		-0.1486	-0.3059		
$CH_3CO_2^-$	-74.12	-74.18	-77	-1.1571	-1.3062	2.87	2.82
F^-	-94.54	-94.85		0.0914	-0.2218		

Table 7: B3LYP/6-31G(2d,p) electrostatics and outlying charge vs. experiment. ΔE_{elect} dc is calculated using double cavity method, ΔE_{elect} dm is calculated using distributed multipoles method. G(expt-calc) is the difference between experimental and calculated free energy (kcal/mol).

Molecules	ΔE_{elect} dc	ΔE_{elect} dm	ΔG expt	OCE dc	OCE dm	G(expt-calc) dc	G(expt-calc) dm
CH_3COOH	-6.48	-6.36		0.0687	0.1210		
$CH_3C(O)CH_3$	-4.87	-4.78	-3.9	0.2230	0.1658	-0.96	-0.88
CCl_4	-1.23	-1.36	-1.1	-0.0480	-0.2327	-0.12	-0.26
CH_4	-0.35	-0.29	2	0.0297	-0.0044	-2.35	-2.28
C_2H_6	-0.34	-0.26	1.8	0.0400	-0.0435	-2.13	-2.06
HC(O)H	-3.64	-3.58	-7	0.1654	0.1737	3.36	3.42
H_2O	-6.43	-6.34	-6.3	-0.0029	0.0604	-0.12	-0.04
HF	-4.92	-4.82	-5.6	0.0350	0.1219	0.68	0.77
CH_3OH	-4.48	-4.39	-5.1	0.0692	0.0732	0.62	0.70
NH_3	-4.79	-4.74	-4.2	0.0037	0.0043	-0.58	-0.53
CH_2CCH_2	-4.44	-4.33		-0.0012	0.0122		
$CH_3CH_2CH_3$	-0.39	-0.32	2	0.0497	-0.0891	-2.38	-2.31
$(CH_3)_3N$	-1.86	-1.81	-3.2	0.0810	-0.0918	1.33	1.38
CH_3^+	-78.72	-78.67		0.9444	0.9698		
NH_4^+	-81.61	-81.43	-77	0.9361	1.0945	-4.61	-4.43
$CH_3NH_3^+$	-72.38	-72.28	-68	1.2459	1.2800	-4.37	-4.27
$(CH_3)_2NH_2^+$	-64.96	-64.90	-61	1.5047	1.4597	-3.96	-3.89
$(CH_3)_3NH^+$	-58.57	-58.58	-57	1.7360	1.5747	-1.56	-1.58
$(CH_3)_3O^+$	-56.51	-56.47		1.6868	1.6082		
K^+	-70.39	-70.62		0.2850	0.0532		
OH^-	-96.36	-96.44	-106	-1.4625	-1.5782	9.64	9.56
CN^-	-74.80	-74.98	-77	-1.3313	-1.5278	2.19	2.01
NO_2^-	-72.94	-73.14	-72	-0.9316	-1.1327	-0.94	-1.13
NH_2^-	-90.13	-90.24	-93	-2.6493	-2.8563	2.86	2.76
BH_4^-	-74.87	-75.21		-4.5404	-5.2531		
BF_4-	-62.30	-62.48		-0.2873	-0.4752		
$CH_3CO_2^-$	-71.38	-71.44	-77	-1.2054	-1.3735	5.61	5.55
F^-	-94.54	-94.85		0.0927	-0.2205		

Table 8: B3LYP/DZV(2d,p) electrostatics and outlying charge vs. experiment. ΔE_{elect} dc is calculated using double cavity method, ΔE_{elect} dm is calculated using distributed multipoles method. G(expt-calc) is the delta between experimental and calculated free energy (kcal/mol).

Molecules	ΔE_{elect} dc	ΔE_{elect} dm	ΔG expt	OCE dc	OCE dm	G(expt-calc) dc	G(expt-calc) dm
CH_3COOH	-7.15	-7.10		-0.0195	-0.0623		
$CH_3C(O)CH_3$	-5.61	-5.57	-3.9	0.1631	0.0248	-1.71	-1.67
CCl_4	-1.10	-1.18	-1.1	-0.0251	-0.1459	-0.01	-0.08
CH_4	-0.53	-0.46	2	0.0255	-0.0383	-2.52	-2.46
C_2H_6	-0.53	-0.44	1.8	0.0390	-0.0635	-2.32	-2.23
HC(O)H	-4.28	-4.27	-7	0.0769	0.0435	2.71	2.73
H_2O	-7.18	-7.17	-6.3	-0.1162	-0.1431	-0.88	-0.87
HF	-5.56	-5.49	-5.6	-0.0220	0.0380	0.04	0.11
CH_3OH	-5.06	-5.03	-5.1	0.0002	-0.0600	0.04	0.07
NH_3	-5.40	-5.46	-4.2	-0.1085	-0.2363	-1.20	-1.25
CH_2CCH_2	-5.64	-5.88		-0.2579	-0.6882		
$CH_3CH_2CH_3$	-0.62	-0.53	2	0.0478	-0.1184	-2.61	-2.53
$(CH_3)_3N$	-2.44	-2.45	-3.2	0.0409	-0.2387	0.75	0.74
CH_3^+	-78.73	-78.67		0.9966	1.0349		
NH_4^+	-81.58	-81.39	-77	1.1621	1.3201	-4.57	-4.39
$CH_3NH_3^+$	-72.33	-72.23	-68	1.4696	1.4866	-4.32	-4.23
$(CH_3)_2NH_2^+$	-64.93	-64.87	-61	1.7504	1.6733	-3.93	-3.87
$(CH_3)_3NH^+$	-58.56	-58.57	-57	1.9935	1.8024	-1.56	-1.57
$(CH_3)_3O^+$	-56.56	-56.51		1.9096	1.8111		
K^+	-70.39	-70.63		0.2742	0.0409		
OH^-	-98.30	-98.57	-106	-2.5272	-2.8571	7.69	7.42
CN^-	-74.59	-74.83	-77	-2.3282	-2.6115	2.41	2.16
NO_2^-	-72.95	-73.19	-72	-1.6529	-1.9148	-0.94	-1.19
NH_2^-	-92.87	-93.33	-93	-3.9993	-4.6217	0.12	-0.32
BH_4^-	-75.17	-75.92		-6.2990	-7.8057		
BF_4^-	-62.07	-62.30		-1.0047	-1.2494		
$CH_3CO_2^-$	-72.97	-73.24	-77	-2.1895	-2.6059	4.02	3.76
F^-	-94.54	-94.85		-0.4591	-0.7731		

Table 9: MP2/6-31G(2d,p) - Electrostatics and outlying charge vs. experiment. ΔE_{elect} dc is calculated using double cavity method, ΔE_{elect} dm is calculated using distributed multipoles method. G(expt-calc) is the delta between experimental and calculated free energy (kcal/mol).

Molecules	ΔE_{elect} dc	ΔE_{elect} dm	ΔG expt	OCE dc	OCE dm	G(expt-calc) dc	G(expt-calc) dm
CH_3COOH	-6.37	-6.26		0.0442	0.0763		
$CH_3C(O)CH_3$	-4.37	-4.29	-3.9	0.2020	0.1273	-0.46	-0.39
CCl_4	-1.21	-1.35	-1.1	-0.0518	-0.2466	-0.11	-0.25
CH_4	-0.39	-0.32	2	0.0270	-0.0138	-2.38	-2.32
C_2H_6	-0.39	-0.31	1.8	0.0366	-0.0575	-2.18	-2.11
HC(O)H	-3.62	-3.43	-7	0.1470	0.1444	3.38	3.57
H_2O	-6.77	-6.71	-6.3	-0.0319	0.0040	-0.46	-0.41
HF	-5.04	-4.96	-5.6	0.0137	0.0893	0.56	0.64
CH_3OH	-4.76	-4.69	-5.1	0.0508	0.0407	0.33	0.40
NH_3	-5.08	-5.05	-4.2	-0.0297	-0.0558	-0.87	-0.85
CH_2CCH_2	-4.76	-4.68		-0.0179	-0.0320		
$CH_3CH_2CH_3$	-0.46	-0.39	2	0.0454	-0.1048	-2.45	-2.38
$(CH_3)_3N$	-2.16	-2.12	-3.2	0.0709	-0.1089	1.03	1.08
CH_3^+	-78.85	-78.83		0.9481	0.9485		
NH_4^+	-81.91	-81.75	-77	0.9154	1.0581	-4.91	-4.74
$CH_3NH_3^+$	-72.75	-72.63	-68	1.2194	1.2644	-4.74	-4.63
$(CH_3)_2NH_2^+$	-65.23	-65.17	-61	1.4885	1.4376	-4.23	-4.16
$(CH_3)_3NH^+$	-58.76	-58.82	-57	1.7652	1.5480	-1.75	-1.82
$(CH_3)_3O^+$	-56.72	-56.69		1.6876	1.5864		
K^+	-70.40	-70.62		0.2974	0.0763		
OH^-	-97.35	-97.47	-106	-1.4075	-1.561	8.65	8.53
CN^-	-74.87	-75.06	-77	-1.3388	-1.5424	2.12	1.94
NO_2^-	-73.43	-73.64	-72	-0.9778	-1.2001	-1.42	-1.64
NH_2^-	-91.38	-91.51	-93	-2.6175	-2.8426	1.62	1.49
BH_4^-	-74.81	-75.17		-4.6923	-5.4414		
BF_4^-	-62.52	-62.68		-0.2539	-0.4187		
$CH_3CO_2^-$	-71.87	-71.92	-77	-1.274	-1.4299	5.12	5.07
F^-	-94.54	-94.85		0.0820	-0.2312		

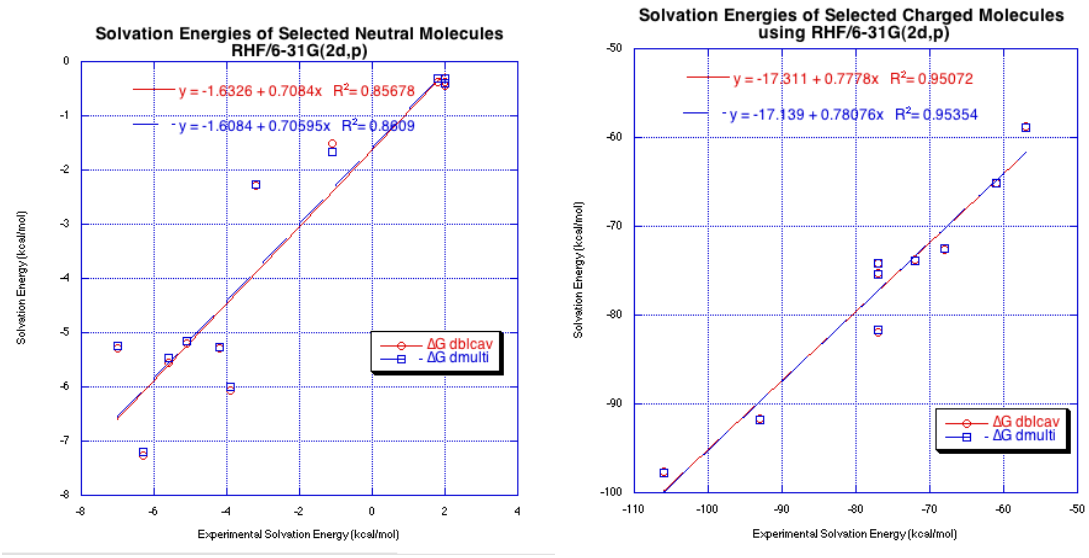


Figure 9: Comparison of Solvation Energies, Experimental vs. Calculated for Neutral and Charged Molecules using RHF/6-31G(2d,p).

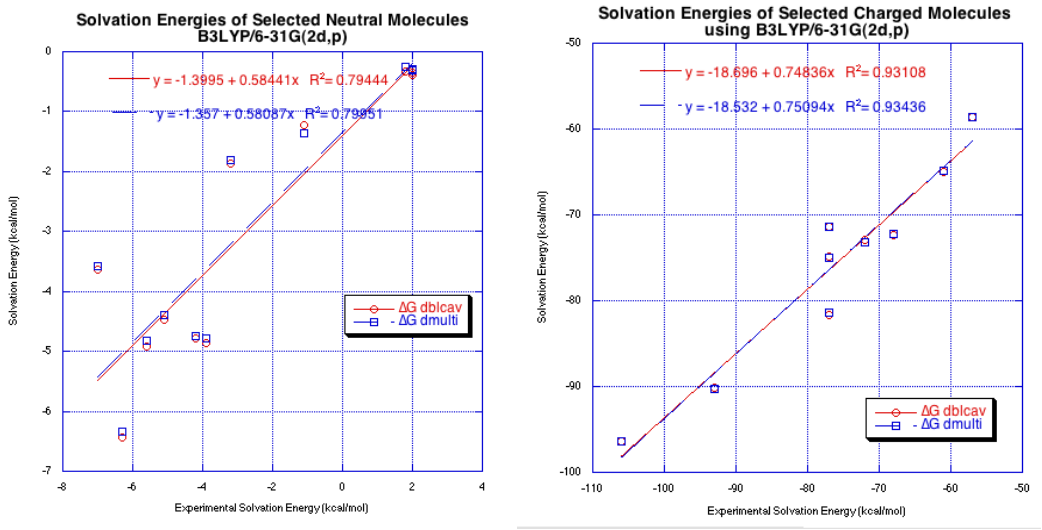


Figure 10: Comparison of Solvation Energies, Experimental vs. Calculated for Neutral and Charged Molecules using B3LYP/6-31G(2d,p).

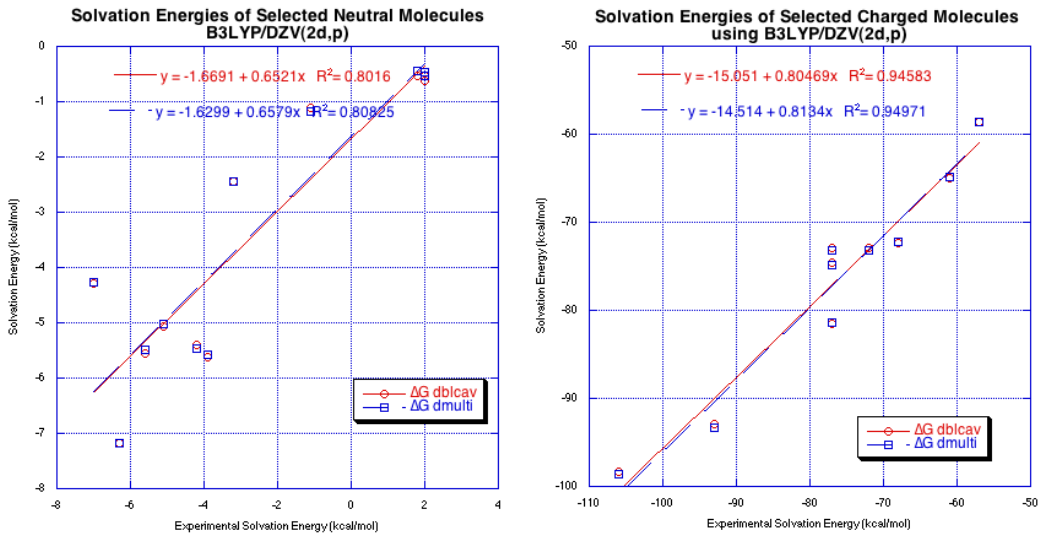


Figure 11: Comparison of Solvation Energies, Experimental vs. Calculated for Neutral and Charged Molecules using B3LYP/DZV(2d,p).

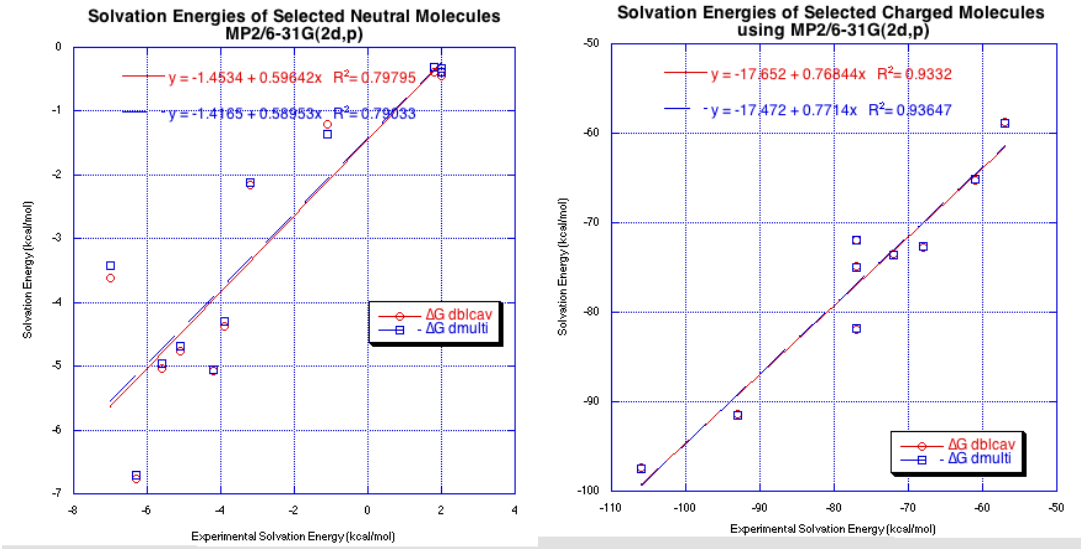


Figure 12: Comparison of Solvation Energies, Experimental vs. Calculated for Neutral and Charged Molecules using MP2/6-31G(2d,p).

Neutral Molecules

Keeping in mind that the calculated ΔE_{elect} up until now involves no direct consideration of non-electrostatic effects. The calculated values still come within 1.56 to 1.69 kcal/mol of the experimental results.

Neutral molecules

Table 6 presents the results using the RHF/6-31G(2d,p) level of theory to calculate the free energy of solvation for the set of neutral molecules with both methods of outlying charge. One first observes that DBLCAV and DMULTI methods

give very similar results. In the case of DBLCAV, the results are between 0.05 kcal/mol (HF) and 2.3 kcal/mol (methane) from the known experimental values. On average, the data are within 1.6 kcal/mol of the respective experimental values. For DMULTI, the results are between 0.04 kcal/mol (CH_3OH) and 2.3 kcal/mol (methane) from experiment, and the average error is 1.5 kcal/mol.

Table 7 presents results using the B3LYP/6-31G(2d,p) level of theory for the same set of molecules. The overall results are very similar, with the average error with respect to experiment of 1.7 kcal/mol for both DBLCAV and DMULTI methods. The results presented in Table 8 shows the results using a slightly different basis set representation, B3LYP/DZV(2d,p), giving an average error with respect to experiment also of 1.7 kcal/mol for both methods.

Table 9 shows the results from the energy of solvation using MP2/6-31G(2d,p). Even with the more conventional wavefunction theory (MP2), the average error with respect to experiment is 1.7 kcal/mol using either DBLCAV method or DMULTI.

Across all wavefunction types considered, the results are only outside the 2 kcal/mol target range with respect to experimental data for methane, ethane and propane, all of which belong to the class of alkanes, which are known to be difficult to evaluate with continuum methods.

The OCE values are also presented across the different levels of theory and methods. Looking at the results, one can observe that the final accuracy in solvation energy is comparable between the two methods, even though the scale and values of their respective OCE can be different. This difference between the two methods can be explained by the way the OCE is calculated for each method (Chapter 3.2.3) and does not reflect the quality of the original method. It is expected, as previously reported,[16,40] that cancellation of error between the chosen level of theory, the solvation method and the cavity construction, play a role in the final result. The energy corrections from the calculations, presented in Table 6, Table 7, Table 8 and Table 9, show a negative energy correction for some molecules (carbon tetrachloride for example) and a positive correction for others (acetone for example). It is related to the shape of the cavity, the wavefunction of the molecule, and the method. That is, depending on the extent and electronic distribution of the wavefunction existing outside the cavity, there will be a commensurate screening charge correction term (and thereby a potential correction, Φ') representing the response of the solvent on

this extra wavefunction. The overall effect of the outlying charge correction will be either represented as an energy gain (negative sign of the OCE) or energy loss (positive sign of the OCE).

From the correlation between calculated and experimental value of solvation energies presented in Figure 9 to Figure 12, the R^2 values are consistent with previously published results and attest to the fact that both method are accurate to the target range of less than 2 kcal/mol error without consideration of nonelectrostatic effects. Additionally, the methods shown here, in particular RHF and DFT, do not have the recently developed dispersion enabled capability, which could explain a certain loss of accuracy for system where dispersion is important.

Anionic and Cationic molecules

The absolute solvation energies for the ionic species is seen to be larger than that for neutral species, as might be expected from the much larger electrostatic interactions of the charged species with the solvent environment that is possible. As might be expected, the discrepancy between calculated and experimental values is larger than what was observed with the neutral species. Table 6 presents the results using RHF/6-31G(2d,p) level of theory to calculate the energy of solvation for the set of cations and anions using both DBLCAV and DMULTI methods. The average error with respect to experiment is 4.1 kcal/mol and 4.0 kcal/mol for DBLCAV and DMULTI, respectively. For DBLCAV, the results are between 1.3 kcal/mol (NH_2^-) and 8.3 kcal/mol (OH^-) of experiment. In the case of DMULTI, the results are between 1.1 kcal/mol and 8.2 kcal/mol of experiment.

Using the B3LYP/6-31G(2d,p), the average error is 4.7 kcal/mol for the DBLCAV method and 4.6 kcal/mol for DMULTI. Unlike the class of neutral molecules, however, using the DZV(2d,p) basis set (Figure 8) gives average errors with respect to experimental data of 3.9 kcal/mol for the DBLCAV method and 3.8 kcal/mol for DMULTI, which is a bigger improvement.

Finally, Table 9 presents the results using MP2/6-31G(2d,p). The average error from experimental values is 4.4 kcal/mol for both DBLCAV and DMULTI methods. In all cases, the largest error occurs with the prediction of the OH^- anion, with more than 5 kcal/mol in error with respect to experiment, however, several experimental values are unknown and so a general trend cannot be drawn here.

An important factor in determination of solvation phenomenon especially for charged species is potential outlying charge error. Results seen above are relatively

typical for continuum models,[10] and can be significantly worse if the model does not include effects of outlying charge. In the present COSab model, the OCE is minimal due to the special methods of correction directly in the algorithm. For cationic species, the difference between calculated and experimental is actually comparable to neutral molecules. All of the OCE corrections are positive for cations, whatever the level of theory is, as expected from the radial symmetric case (Chapter 3.3.3). The magnitude is comparable to neutral molecules' OCE as the results of their small outlying charge error. However, the effects of OCE for cations can still be important, and should be accounted for. The OCE for the anionic species is more extensive due to its more diffuse negative charge density that extends beyond the atomic coordinate-generated cavity, unlike the cation species. The magnitude of the OCE is therefore larger than in cations and neutral molecules, reaching up to 5% of the calculated solvation energy as outlying charge.[40]

As in previously published work,[16] the prediction of free energy of solvation of small molecules is quite good, even considering the missing nonelectrostatic component. The additional OCE methods in the implementation of COSab also increase the accuracy of the model.

3.4.3 Effect of Basis Set

In this section, we show the relative behavior of the two models with respect to basis set quality. Table 10 summarizes structure and dipole results for both gas phase and solution phase calculations of H_2O at a variety of basis sets using the RHF wavefunction method. Figure 13 shows the general trends of basis set on the O-H bond distance, and the predicted dipole moment in H_2O for both gas phase and solution phase.

Table 10: Basis set effects on structure and properties of H_2O. Experimental value of dipole of water in gas phase: 1.85 D.[60] RHF level of theory.

Basis Set	R(OH) Å gas phase	R(OH) Å dblcav	R(OH) Å dmulti	μ (Debye) gas phase	μ (Debye) dblcav	μ (Debye) dmulti
STO-3G	0.9894	0.9870	0.9870	1.71	1.83	1.83
3-21G	0.9667	0.9682	0.9681	2.39	2.65	2.65
6-31G	0.9497	0.9524	0.9523	2.50	2.81	2.81
6-31G(d)	0.9473	0.9503	0.9503	2.20	2.52	2.52
6-31G(d,p)	0.9431	0.9460	0.9460	2.15	2.47	2.47
6-31G(2d,p)	0.9414	0.9442	0.9442	1.96	2.31	2.30
6-311G(2d,p)	0.9423	0.9451	0.9451	2.06	2.40	2.40
DZV(d)	0.9468	0.9500	0.9500	2.26	2.60	2.60
TZV(d)	0.9431	0.9454	0.9454	2.36	2.72	2.72
TZV(2d,p)	0.9412	0.9443	0.9444	2.13	2.49	2.49
cc-pVDZ	0.9464	0.9494	0.9494	2.05	2.35	2.35
Expt	0.957			2.3		

The predicted values obtained for both methods are in good agreement with the experimental values. The value of 2.3 D for the experimental value of the dipole of water has been taken in solution.[60] Results would also indicate that even results generated using the RHF wavefunction type, RHF/6-31G(2d,p), are reliable. The value of dipole of water in solution (2.3) is in agreement with experimental values estimated at 2.3-2.4 D.[60] The value in gas phase with the same basis set is also accurate compare to experimental result (1.96 D vs 1.85 D for calculated and experimental, respectively).

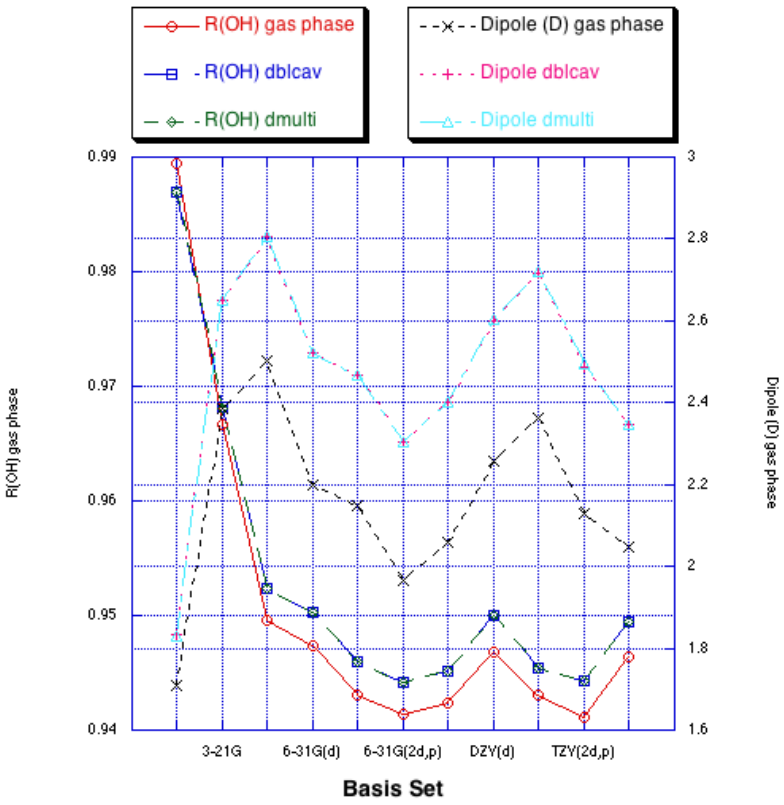


Figure 13: Effects of basis set on bond distance and dipole moment for H_2O and solution phase.

Figure 13 shows variations in OCE for H_2O and OH^- using a variety of basis sets. As previously reported,[16] there is a strong dependency of OCE with the choice of basis set, especially in the case of anions. It has also been reported[40] that using very diffuse basis set or triple zeta, as one would use for anions, the OCE of neutral and anions can be as much as 20% of the dielectric energy, which is the total energy gain due to the interaction of the solute with the continuum. The observed changes in bond length and dipole in gas phase and solution phase (using both methods) follow the same trend (Figure 13), even though one must assume that OCE is not the only variable that is modified by the different basis sets. This basis set study illustrates the importance of wavefunction and basis set for obtaining accurate results for energetics, properties and structure.

Table 11 presents the variation in solvation energy and outlying charge effect as a function of basis set. In particular, a significant difference is observed between double and triple zeta basis set representations, for capturing all of the OCE. Moreover, one can see the importance of adequate basis set size for capturing the

OCE for anions. While the maximum OCE for H_2O across all basis sets (except minimal basis set) is no more than ½ kcal, the OCE for OH- can be as much as 5-6 kcal. We note that the DMULTI method typically shows slightly larger OCE errors than the DBLCAV method. For quite diffuse molecules, this would be the case since the DMULTI method is able to more adequately capture the extended wavefunction through secondary representation in terms of a distributed multipole inside the cavity. In contrast, the DBLCAV method is based on a parametric representation of a secondary cavity at 1 Å from the molecular cavity, and if the wavefunction extends beyond this secondary cavity, all of the OCE can not be assessed.

Table 11: Basis set effect on solvation energy and OCE for H_2O (experimental value: 6.31 kcal/mol) and OH^-.

	H_2O					OH^-				
Basis Set	E gas (a.u.)	phase	ΔE_{elect} dc kcal/mol	OCE dc	ΔE_{elect} dm kcal/mol	OCE dm	E gas (a.u.)	phase	ΔE_{elect} dc kcal/mol	OCE dc	ΔE_{elect} dm kcal/mol	OCE dm
STO-3G	-74.96		-4.44	0.1221	-4.31	0.2345	-74.05		-92.89	-0.5543	-92.99	-0.6723
3-21G	-75.58		-9.03	0.0461	-8.92	0.1454	-74.86		-96.59	-0.4323	-96.65	-0.5018
6-31G	-75.98		-10.71	-0.0938	-10.67	-0.0715	-75.31		-98.99	-1.2572	-99.15	-1.4387
6-31G(d)	-76.01		-8.36	-0.0856	-8.33	-0.0741	-75.33		-98.23	-1.3346	-98.36	-1.4972
6-31G(d,p)	-76.02		-8.08	-0.0563	-8.03	-0.0313	-75.33		-97.99	-1.3498	-98.11	-1.5011
6-31G(2d,p)	-76.03		-7.25	-0.0506	-7.21	-0.0268	-75.33		-97.69	-1.3446	-97.82	-1.5058
6-311G(2d,p)	-76.05		-7.68	-0.0703	-7.65	-0.0585	-75.36		-98.17	-1.6475	-98.32	-1.8260
DZV(d)	-76.04		-9.01	-0.1981	-9.05	-0.2769	-75.37		-99.89	-2.5482	-100.20	-2.9218
TZV(d)	-76.04		-10.04	-0.2924	-10.16	-0.4650	-75.39		-101.29	-4.5404	-101.92	-5.3079
TZV(2d,p)	-76.06		-8.47	-0.2758	-8.61	-0.4573	-75.40		-100.36	-4.5458	-100.97	-5.2955
cc-pVDZ	-76.03		-7.18	-0.0386	-7.14	-0.0141	-75.33		-97.63	-1.1525	-97.77	-1.3145

3.5 Beyond a Perfect Conductor Hypothesis: COSab+RS and COSab-D

3.5.1 COSab + RS

With COSab, we properly describe the electrostatic influence of solvent on solute molecules. For a given electrostatic field of the solute, the dielectric medium screens the field by a polarization of the medium, which is represented by surface charges on the boundary surface between the solute and solvent environment. The method as implemented has been shown to be an efficient and accurate method for prediction of electrostatic solvation phenomenon. However, a missing component of the model is the treatment of first solvation shell phenomenon – so-called non-electrostatic effects.

In accord with the discussion in Section 3.2.2, the COSab model is based on the idea that the medium is a conductor. As such, there is perfect matching of the surface charge interface segments between the solute and the solvent. However, non-ideality of solvent environment (i.e., the non-conductor-like behavior) would lead to a non-ideal pairing between surface segments for many solute molecules as well as between solvent molecules. Treatment of such non-ideality can be carried out based on a model by Klamt, the COSMO-RS (Conductor-like Screening Model for Real Solvent) framework.[39,42,61] In this model, the deviation from ideal screening is described as pairwise misfit interactions of the idea screening charges on individual contact pieces of the surface boundaries between the solvent and the solute, or between the solvent molecules. In this work, two strategies for including effects of non-electrostatics are implemented in GAMESS, and will be discussed here.

Within the underlying COSMO-RS model, the solvent environment is represented as a dense packing of molecules, and this environment is broken down to an ensemble of pair wise interacting surface segments, just as discussed earlier for the molecule itself. The overall solvent interaction is represented by surface descriptors, the most important of which is the screening charge density, that is, the screening charge per segment area ($\sigma_t = q_t/a_t$). These screening charges are already calculated from the basic COSab calculation. Therefore, from these results, we have an efficient way to calculate the total energy of the artificial ensemble of solvent molecules, by

simply performing a COSab calculation for each different type of solvent molecule in the ensemble.

If the screening charges of two interacting surface segments σ and σ' differs, the interaction energy will include a "misfit" term, from the interaction of the associated segment pairings, as represented on Figure 14, in accord with Equation 19 where a_{eff} is the effective contact area.

$$E_{misfit}(\sigma,\sigma') = a_{eff}\frac{\alpha'}{2}(\sigma+\sigma')^2 \qquad \textbf{19}$$

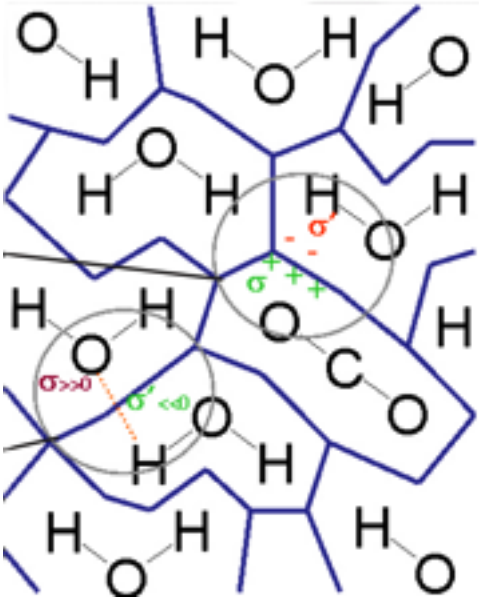


Figure 14: Schematic construction of molecular cavities and misfit of charges.[57]

Hydrogen bonds are important interactions in condensed media not specifically accounted for in the basic COSab electrostatic treatment. Such directed interactions, occurring in the first solvation shell, need to be accounted for as another descriptor type. The COSMO-RS methodology assumes that a hydrogen bond is formed between a sufficiently polar piece of surface of a donor and an acceptor, and that the bond strength is dependent on the actual polarity of these pieces. The H-bond descriptor is determined via the relationship.

$$E_{HB} = a_{eff}c_{HB}\min(0;\min(0;\sigma_{donor}+\sigma_{HB})\max(0;\sigma_{acceptor}-\sigma_{HB})) \qquad \textbf{20}$$

Equations 19 and 20 contain five parameters. First, there is an interaction parameter α' (α'=1385 kJ nm^2/mol/e^2). This parameter has been evaluated from the

group contribution model UNIFAC, where a statistical thermodynamics approximation of pair-wise interacting surfaces pieces is fit to experimental data.[62] Equation 20 assumes that a pair of independent surface segments can describe residual non-steric interactions. The effective contact area, $a_{eff,}$ is the thermodynamically independent contact area between two surface segments (a_{eff} = 0.0767 nm^2). Hydrogen bonding is described by two screening charge densities, the hydrogen bond strength c_{HB}, and the threshold for hydrogen bonding σ_{HB} (σ_{HB} = 0.79e/nm^2). It is generally considered that a hydrogen bond will be present when two sufficiently polar pieces of surface of opposite polarity are in contact (± 1 e/nm^2).

In addition to electrostatic misfit and hydrogen bond interactions, COSMO-RS also takes into account van der Waals (vdW) interactions between surface segments using Equation 21, where τ_{vdW} and τ'_{vdW} are adjustable parameters. The van der Waals energy is dependent only on the element type involved in the surface contact. Klamt et al. has provided element-dependent optimized parameters for these.[42] These element specific parameters τ have been fitted to experimental data.[62]

$$E_{vdW} = ae_{ff}(\tau_{vdW} + \tau'_{vdW}) \qquad \textbf{21}$$

Since all interactions of the surface are ultimately determined by the element type and the local polarization charge densities σ, only the net composition of the surface of a molecule X with respect to σ is important for the statistical thermodynamics of local pairwise surface interactions. Therefore, averaging these surface charge densities over contact segments can be carried out with a distribution function, or '*sigma profile,* $p^X(\sigma)$, which gives the probability of finding a mean screening charge density, σ, on a contact segment of a molecule X. The σ-profile for the entire molecule of interest $p_S(\sigma)$ can be expressed as the sum of all components weighted by their mole fraction, according to Equation 22.

$$p_S(\sigma) = \sum_{i \in S} x_i p^{X_i}(\sigma) \qquad \textbf{22}$$

Using the σ-profile, the chemical potential of the surface segments can be calculated by solving a coupled set of non-linear equations.

$$\mu_S(\sigma;T) = -\frac{RT}{a_{eff}}\ln\left[\int p_s(\sigma')\exp\left\{\frac{a_{eff}}{RT}\left[\mu_S(\sigma';T) - E_{misfit}(\sigma,\sigma') - E_{HB}(\sigma,\sigma')\right]\right\}d\sigma'\right] \quad \mathbf{23}$$

The chemical potential $\mu'_s(\sigma)$ of a surface segment with screening charge density σ in an ensemble described by normalized distribution function $p_s(\sigma)$ is a measure of the affinity of the system S to a surface of polarity σ. This is formally called the σ-potential of the solvent.

Based on COSab calculations for molecules embedded in a dielectric, GAMESS can now generate a file containing a description of the cavity and the screening charge. A post-treatment using the COSMOtherm software with the GAMESS file enables a COSMO-RS calculation to be carried out. With the same software, one can additionally generate a σ-profile for the molecule.

3.5.2 The Self Consistent DCOSMO-RS Approach

The self-consistent calculation is a pure electrostatic model, and the post-treatment for non-electrostatic (COSab-RS) is not fully self-consistent. In other words, the non-electrostatic effect is not directly integrated in the SCF of the QM procedure. Integrating the COSMO-RS model into the iterative procedure of the ab initio SCF calculation is made possible through the use of the chemical potential that must be calculated initially, as described and shown in equation 24 above.

$$E' = E(\Psi^{solv}) + E^{RS} \quad \mathbf{24}$$

$$E^{RS} = E^{diel} + f_{pol}\sum_t a_t\mu_S(\sigma) + \mu_{C,S}^{Xi} + kT\ln x_{Xi} \quad \mathbf{25}$$

In Equation 25, the factor $\int_{pol}$ accounts for the solute-solvent back polarization. The default value is 1 for now, but a finer tuning is possible. The σ-potential, μ_S, depends on the temperature, the composition of the solvent, and the COSMO-RS parameter set, which needs to have been previously calculated from a COSab calculation.[63]

The solvation influence within the context of the COSMO-RS model can be viewed as a correction of the ideal screening charges appearing in a basic COSab

conductor-like screen model. To obtain a representation of the σ-potential $\mu_S(\sigma)$, we use a linear combination of six polynomials, with exponents I,[62] as shown in Equation 26,

$$f_i(\sigma) = \sigma^i \text{ for i}\geq 0 \qquad \textbf{26}$$

together with two functions that describe the near linear part of the σ-potentials in the hydrogen bond region (otherwise these regions would cause poor convergence in a purely polynomial Taylor series):

$$f_{-2/-1}(\sigma) = f_{acc/don}(\sigma) \cong \begin{Bmatrix} 0 \;\; if \;\; \pm\sigma < \sigma_{HB}^{tresh} \\ \pm\sigma - \sigma_{HB} \;\; if \;\; \pm\sigma > \sigma_{HB}^{tresh} \end{Bmatrix} \qquad \textbf{27}$$

A hydrogen bond threshold σ_{HB}^{tresh} is used, as well as a parabolic smoothing of the edge of these functions. Consideration of the large number of different solvents leads to the use of Taylor-like expansion of the form shown in Equation 28. Using and expansion with m=3 is usually sufficient.

$$\mu_S(\sigma) \cong \sum_{i=-2}^{m} c_S^i f_i(\sigma) \qquad \textbf{28}$$

The contribution of the second moment is better described using the dielectric energy since it is a measure of the overall polarity of the solute. The method neglect this second term and account for this contribution by scaling the COSMO charge and dielectric energy with the following factor:

$$f_{crs} = 1 - c_2 \frac{f_{el}}{c_{up}} \qquad \textbf{29}$$

where, c_{up} denotes a typical c_2 value of an unpolar substance, like hexane, and f_{el} is the screening factor $f(\varepsilon)$ for the electronic polarization ($\varepsilon = n^2$) of a typical organic substance ($n\approx 2$). The parameter, f_{crs}, is used to scale the COSab solute potential.

A DCOSab-RS self-consistent field cycle follows the following scheme:

1. The COSab screening charges, q_i, and the corresponding dielectric energy, E_{diel}, are calculated from the solute potential.(Equation 6)
2. The charges, q_i, are used to calculate the total energy E' and the COSab-RS potential, ϕ^{RS}, which in turn is used to determine the q^{RS}_t.
3. The full operator matrix is added to the Fock/Kohn-Sham matrix.

At the completion of a DCOSab-RS calculation, an outlying charge correction of the energy is calculated as the difference of the E^{RS} energy obtained from the uncorrected and the corrected COSMO charges q_i.

3.5.3 Non-electrostatic Effect Contributions

From our previous work and other literature studies,[16,62] as well as calculations presented in this work, we find that the electrostatic free energy of solvation can, in general, be calculated within an error of 1-2 kcal/mol, depending on the quality of the wavefunction and the basis set. While the electrostatic interaction is the dominant contribution to the free energy of solvation, particularly in polar media, the non-electrostatic contribution is important to consider for a more accurate prediction.

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nel} \qquad \textbf{30}$$

$$\Delta G_{nel} = \Delta G_{cav} + \Delta G_{dr} \qquad \textbf{31}$$

The primary non-electrostatic contributions are generally broken down into contributions from formation of the cavity, ΔG_{cav}, and contributions from dispersion and repulsion, here represented as one term ΔG_{dr} (Equation 31). The cavity formation for chemical reaction in liquid phase is the largest term.[57] The dispersion-repulsion term is the result of the addition of several small contributions, originated from the direct interaction of the reactants with the solvent molecules.

Most of the continuum solvent models that have been proposed employ empirical terms to account for any nonelectrostatic effects.[64,65] In the context of the COSab-based model, results of nonelectrostatic effects can be considered within the COSab-RS method. This extension has been implemented within the SCF calculation in a direct manner in this work, the DCOSab-RS approach. The important data from the nonelectrostatic contribution is described in the generated cosmo file with a set of

coordinate for each segment of the cavity associated with their respective charges and potentials. Using this data, one can subsequently carry out a simple analysis calculation using the COSMOtherm program of Klamt,[57] using the special keyword, Spot[57] in the COSMOtherm input file. This then produces what are known as σ-potentials (Equation 22) for the particular molecule in the solvent environment. The same file contains the σ-potentials coefficient, which can be used in the polynomial in Equation 26. The procedure for carrying out this extended calculation, DCOSab-RS, is illustrated in the flow-diagram shown in Figure 15.

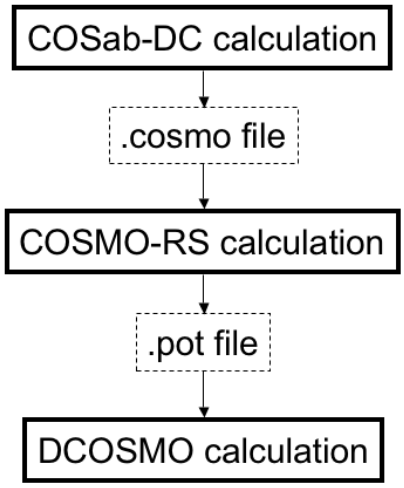


Figure 15: DCOSMO-RS workflow.

3.5.4 Sigma Profiles

In chapter 3.5, we saw the way COSMO-RS describes molecular interactions based on the view of pair wise interacting surface segments. The difference between the screening charge densities σ and σ' of a contact pair is a measure for the misfit of the screening charge density on both segments in the real segment using Equation 20, where a_{eff} is the effective contact area between two surface segment. The basic assumption made with Equation 19 is that pairs of geometrical independent surface segments can describe non-steric interactions. The surface segment, a_{eff}, has to be chosen to effectively correspond to a thermodynamically independent entity. Since there is no easy way to define a_{eff} using first principles, it needs to be an adjustable parameter. The value used in the COSMO-RS method has been chosen based on DFT/COSMO calculations carried out with the Becke-Perdew (BP)[59,66] functional

together with a triple-ζ valence plus polarization (TZVP) basis set.[57] The value used in COSMOtherm is 0.0767nm^2.

The polarization charge densities, σ_I, on the molecular surface are the only necessary properties for evaluation of surface interaction energies. GAMESS has been enabled to generate .cosmo files from a COSab calculation using a new keyword, COSWRT, that was programmed in the $COSGMS group. The .cosmo file contains the charge density information, q_i/s_i, for each surface segment, σ_i. The surface segments have area on the order of 5.10^{-3} nm^2 to 5.10^{-5} nm^2, and the effective contact surface has been defined at a_{eff}=0.0767 nm^2. However, the calculated COSab charge densities are more localized. The link between the microscopic surface interaction energies and the macroscopic thermodynamic properties of a liquid demands a statistical averaging in the ensemble of interacting surface segments. Thus, values of σ are averaged over larger areas using a Boltzmann distribution.

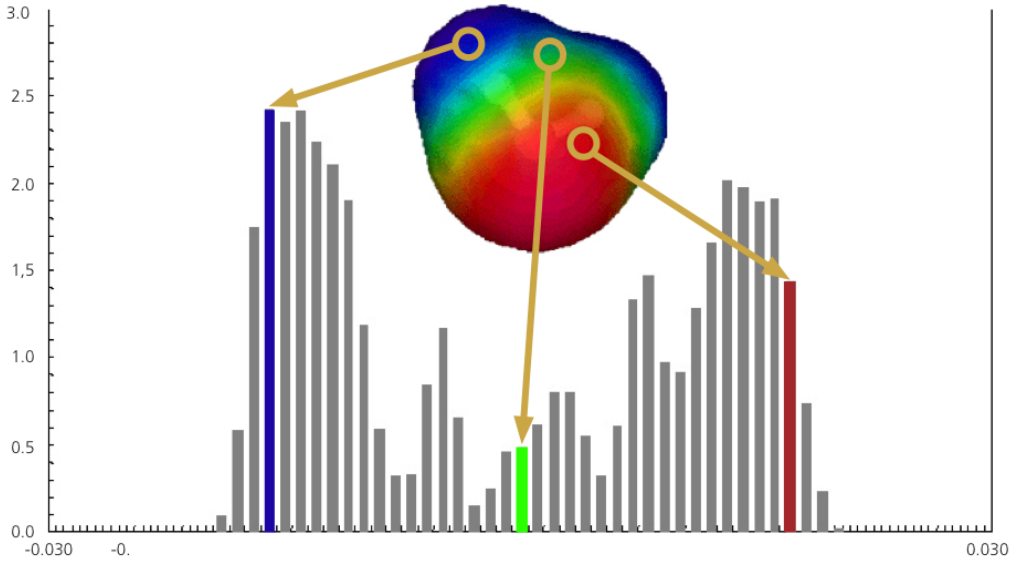


Figure 16: sigma profile of water.

To describe the composition of the surface segment ensemble with respect to the solvent interactions, only the probability distribution of σ really needs to be known. The net composition of the surface of a molecule with respect to σ can then be used to interpret the variation in interaction characteristics of the molecule with the solvent. In this way, the full 3D set of σ information for the molecular surface is reduced to a histogram representing the probability distribution $p^X(\sigma)$. Such a graphic representation is called a "σ-profile", and shows a breakdown of the surface as a

function of expected polarity in specific intervals, [σ-dσ/2,σ+dσ/2] (e.g., see Equation 22).

An example of a σ-profile for water is shown in Figure 16. The entire σ-profile of water spans the range of ±2 e/nm^2, as is the case for most organic molecules. We can see two major peaks, originating from the negative polar regions of the electron lone-pairs of the oxygen atom, and from the positive polar hydrogen atoms. One must be careful in the interpretation of the sign in such a plot. There is an inversion of the sign for polarization charge density,σ, as compared to the molecular polarity. The peak from the negative lone-pair is displayed in red on the right side at positive 1.5 e/nm^2, while the peak from the positively charged hydrogen is located on the left side at negative -1.5 e/nm^2. The hydrogen bond threshold is evaluated at ±0.79 e/nm^2 by fitting to experimental data.[63] Both peaks in the water sigma profile are located beyond this threshold value, an indication that water molecules are able to form hydrogen bonds. The σ-profile of water explains its unique property to enable very favorable pairing of positives and negatives surfaces and formation of hydrogen bonds.

The application of σ-profiles is more generally illustrated in Figure 18, with a selection of common organic solvents. Peaks on the positive side of the graph occur for example, for compounds containing electronegative lone-pairs (e.g., water, methanol, and acetone (propanone)). Again, water also has a peak on the positive side with its hydrogens, as well as the methyl group in methanol. In the central part of the graph around zero or very small positive and negative values of σ, one finds the apolar regions of molecules, e.g., within benzene or toluene.

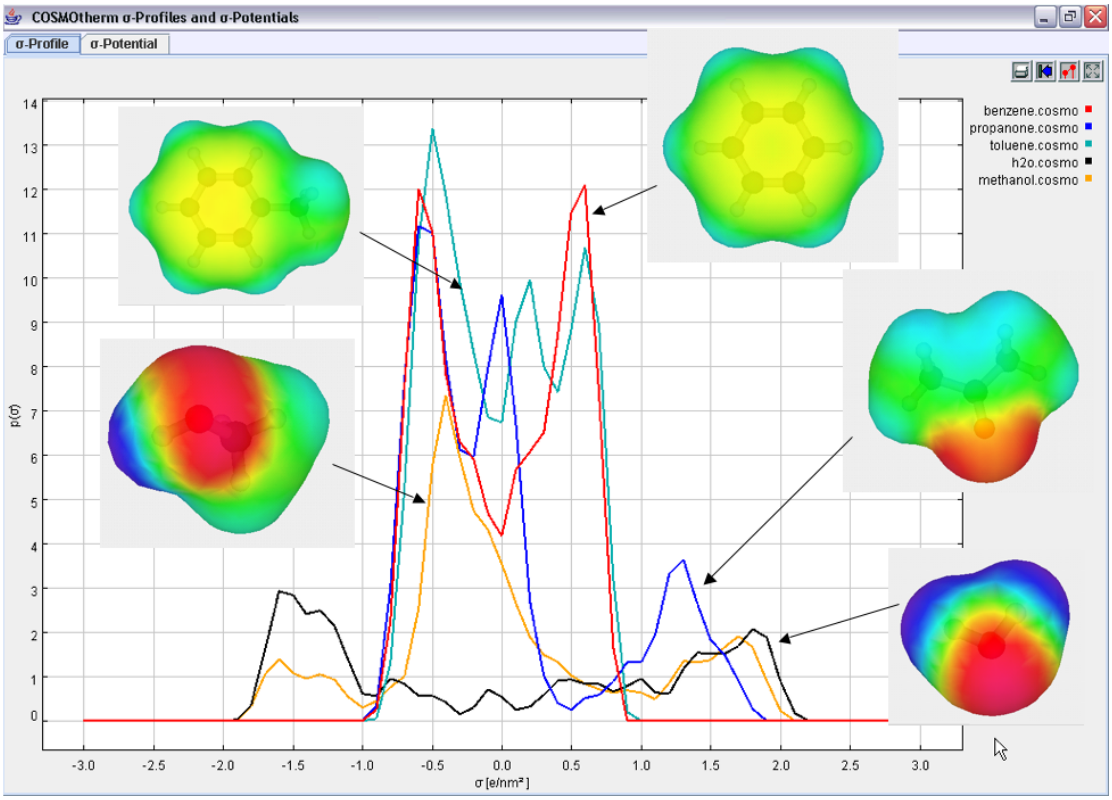


Figure 17: Sigma-profiles of common solvent molecules.

The 2D information of a σ surface as visualized in Figure 17, can also be represented in 3D via a virtual reality markup-up language (VRML) file. Such files are also generated from the .cosmo file output of a solvation run within GAMESS. These 3D colored molecular surface pictures for each molecule are displayed on the graph of Figure 17.

Since σ-profiles describe the polarity of the molecular surface, they can be very useful for understanding molecular reactivity. For example, in Figure 17, the σ-profile of benzene shows two well-separated peaks at σ = +0.5 and at σ = +-0.5, corresponding respectively to the π-face and the hydrogen-belt of the benzene. Toluene shows an additional peak at σ=0, due to the methyl group. Additionally, as compared to benzene, the negative side of the toluene peak is larger. We note that a simple dielectric model would not describe the difference between benzene and toluene, since both have near identical dielectric constants (ε=2.247 and ε=2.379 respectively).

Unlike water, with its broad and symmetric σ-profile with two pronounced peaks, σ=±1.5, from the polar lone-pairs of oxygen and the two polar hydrogens, methanol has only one donor hydrogen peak at -1.5; the oxygen peak is very similar

to water. In addition, the methyl group is shifted to the left due to the polarization from the hydroxyl group.

Acetone (here noted with its IUPAC name: propanone) has a very asymmetric σ-profile. The carbonyl oxygen peak is different from that of water or methanol. There are no donors, and the negative polarity of the oxygen is compensated by the large polarized alkane structure of the two methyl groups. The asymmetry observed in the σ-profile, and the strong electrostatic misfit contribution, underlies the relatively high pressure observed for acetone.

In general, analysis of chemical reactivity can be greatly facilitated through the use of σ-profiles, which provide a detailed quantitative description of the polarity and hydrogen bonding features of solutes and solvents.

3.6 General Application of Models

3.6.1 Predicting Small Molecule Solvation Free Energies

In this section, solvation properties are determined for a large variety of functionalities, to compare the performance and predictability of the solvent methods in GAMESS. The data set is composed of 217 molecules, with a variety of different functionalities, including primarily molecules with elements H, C, N, O and Cl. The data set has been previously determined as a benchmark database,[42] where reliable experimental data is available, and the size of the molecular systems is reasonable for computational efficiency of testing. Experimental data were taken from a variety sources.[67,42,68] Comparison between theoretical and experimental results is also measured using ΔG_{hydr} values, the equivalent to Henry's law constants for the water/air system.[9]

DFT is known to be a reliable theory for accurate ground-state properties, especially ground-state charge distributions.[55] For benchmarking, gas phase optimization geometries are carried out using the BP86/TZVP level of theory, as a preferred method for such computations.[42] Moreover, the necessary COSMO-RS parameters for nonelectrostatic contributions are determined using the TZVP basis set.

All structures have been optimized in gas phase, and the geometries used for solvent phase calculations (ε=78.4, NSPA=92 and COSRAD=1.3), with the three available methods: COSab-DC, COSab-DM and DCOSab-RS, all using BP86/TZVP.

Results are presented in Annex A, including solvation energies of every molecule, outlying charge correction, dielectric energy, and sum of polarization charge, for each of the three methods. Extracted results are also presented in Figure 18 for COSab-DC, Figure 19 for COSab-DM, and Figure 20 for DCOSab-RS. The graphs show a subset of the database of molecules, with C, H and O atoms classed according to functionality: alkanes, alkenes, alkynes, alcohols, ethers, carbonyl, acids, esters and aryls compounds. Graphs noted 'b' represents the molecules containing nitrogen, classed according to amines, amides, aryls, nitriles and nitro. Finally, graphs designated 'c' represents the molecules containing halogens, in particular chlorine atoms.

If we carry out a correlation analysis for the three different models, comparing experimental and theoretical for the full set of 217 molecules, we obtain the following statistical equations:

COSab-DC: $y = 3.09 + 1.14x$, $R^2=0.906$
COSab-DM: $y = 3.15 + 1.1.5x$, $R^2=0.925$
DCOSMO-RS: $y = 1.32 + 1.00x$, $R^2=0.888$

Looking at these equations, COSab-DC and COSab-DM appear to have the same general behavior, with an error with respect to a perfect fit between experimental and theoretical results of about 3.1 kcal/mol, and a non-linearity around 10-15%. In the case of the DCOSMO-RS method, where nonelectrostatic effects are now included, one sees that a linear behavior is observed. So, while the theoretical results give a small deviation from experimental results across the set of molecules (on the order of 1.32 kcal for this model), the linearity between observed and predicted is ideal. Comparing this error to the other two methods, we see that the nonelectrostatic contribution is both decreasing the overall error of prediction (1.32 vs 3 kcal, with and without nonelectrostatic contributions, respectively), as well as improving the overall behavior of the model.

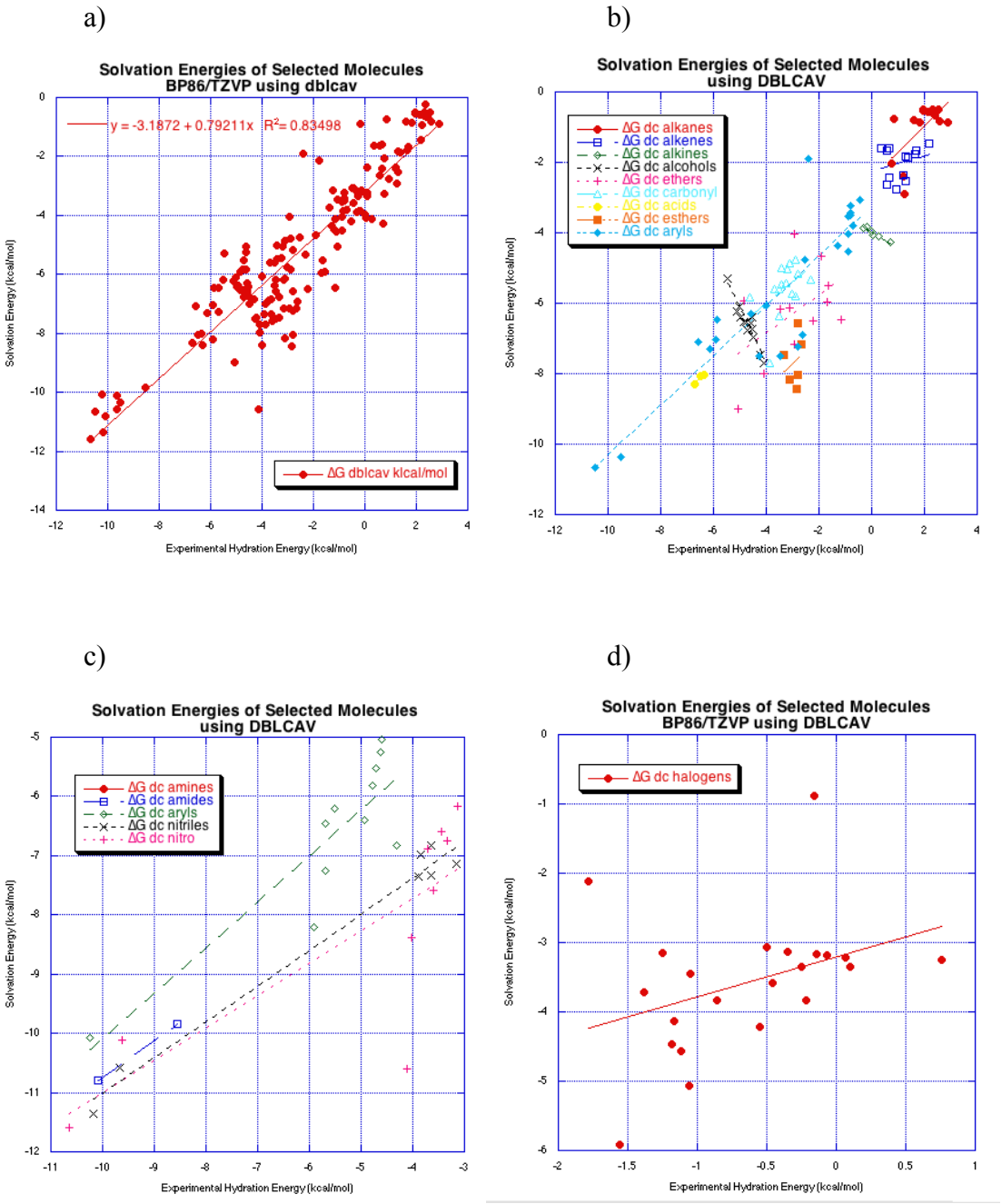
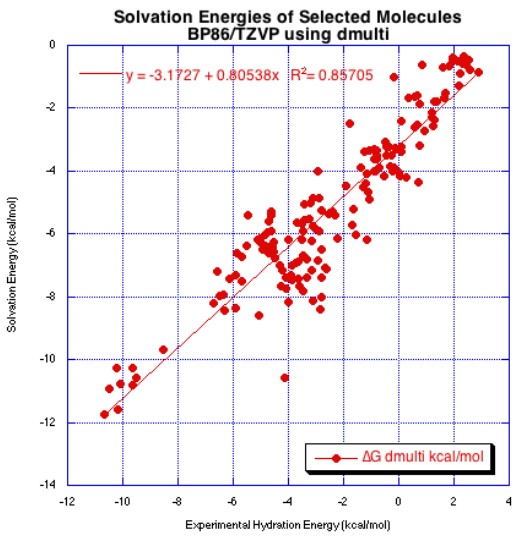
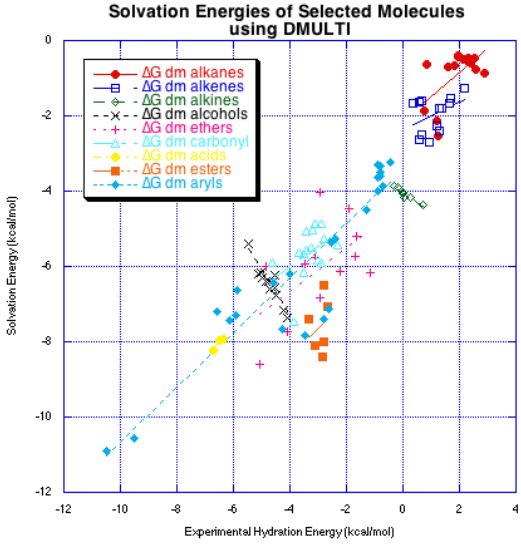


Figure 18: Solvation energies of selected molecules with double cavity method COSab-DC. a) All b) with elements C, H, O c) with element N d) with element Cl.

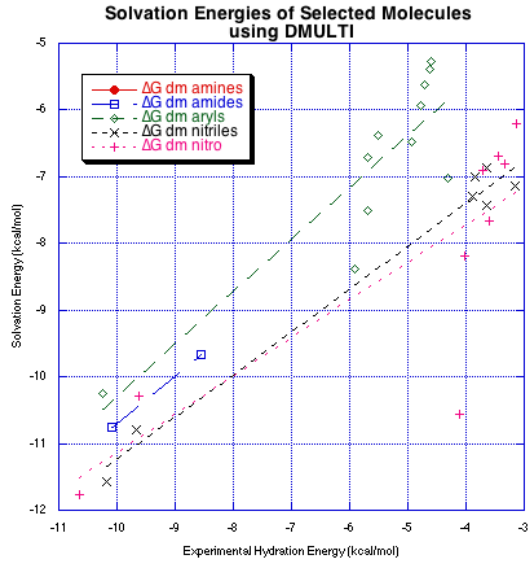
a)



b)



c)



d)

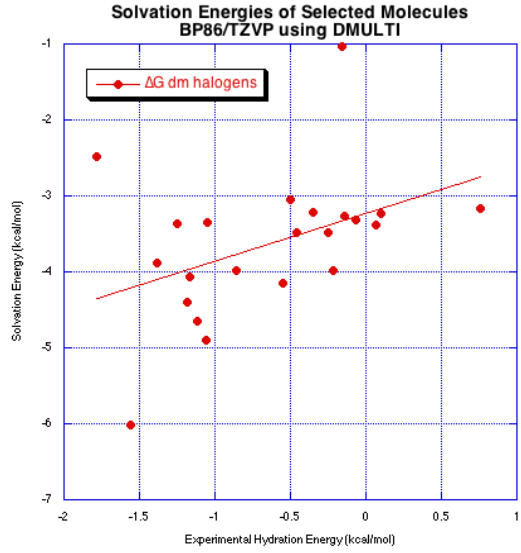


Figure 19: Solvation energies of selected molecules with distributed multipoles method (COSab-DM). a) All b) with elements C, H, O c) with element N d) with element Cl.

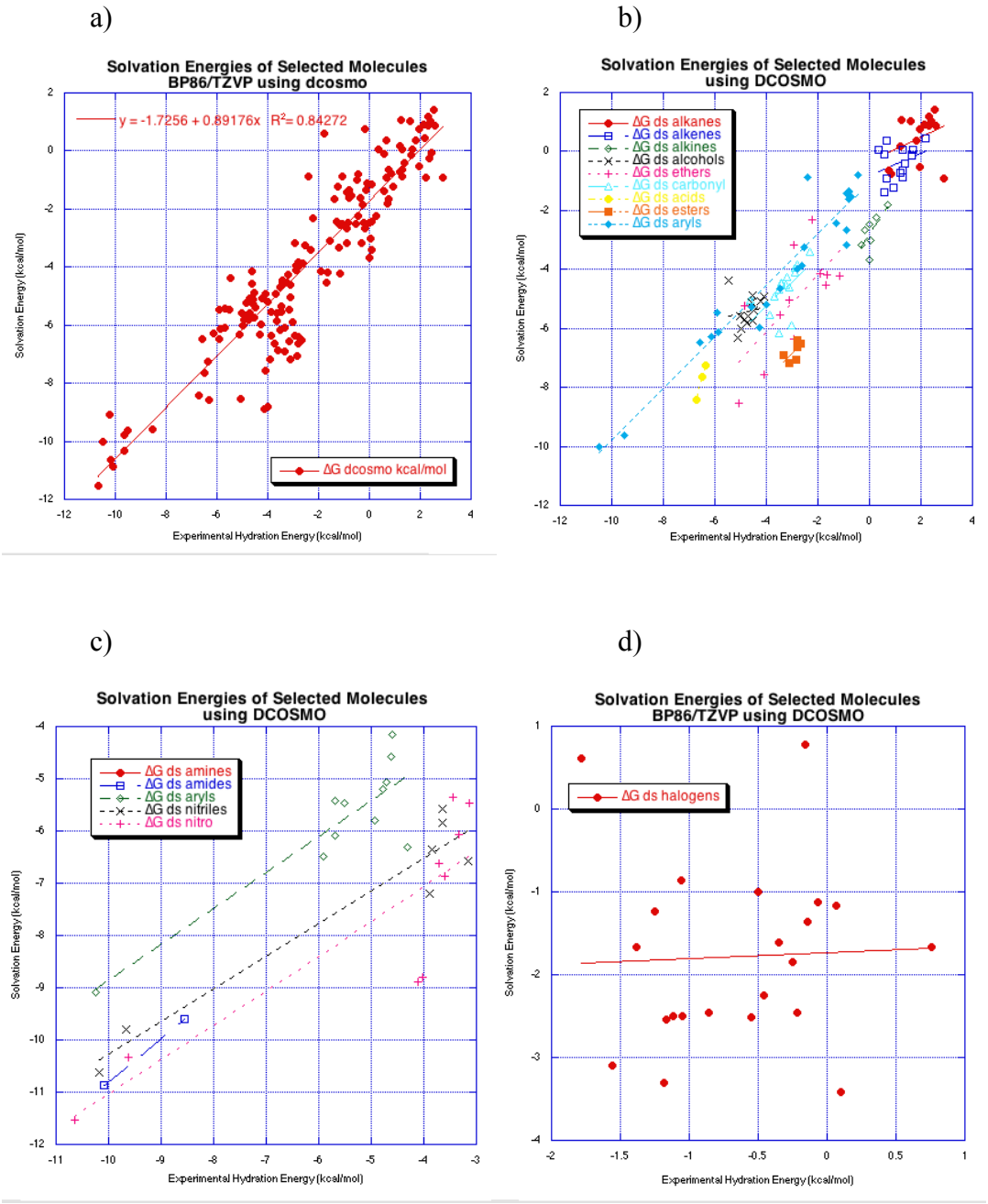


Figure 20: Solvation energies of selected molecules with DCOSMO-RS method. a) All b) with elements C, H, O c) with element N d) with element Cl.

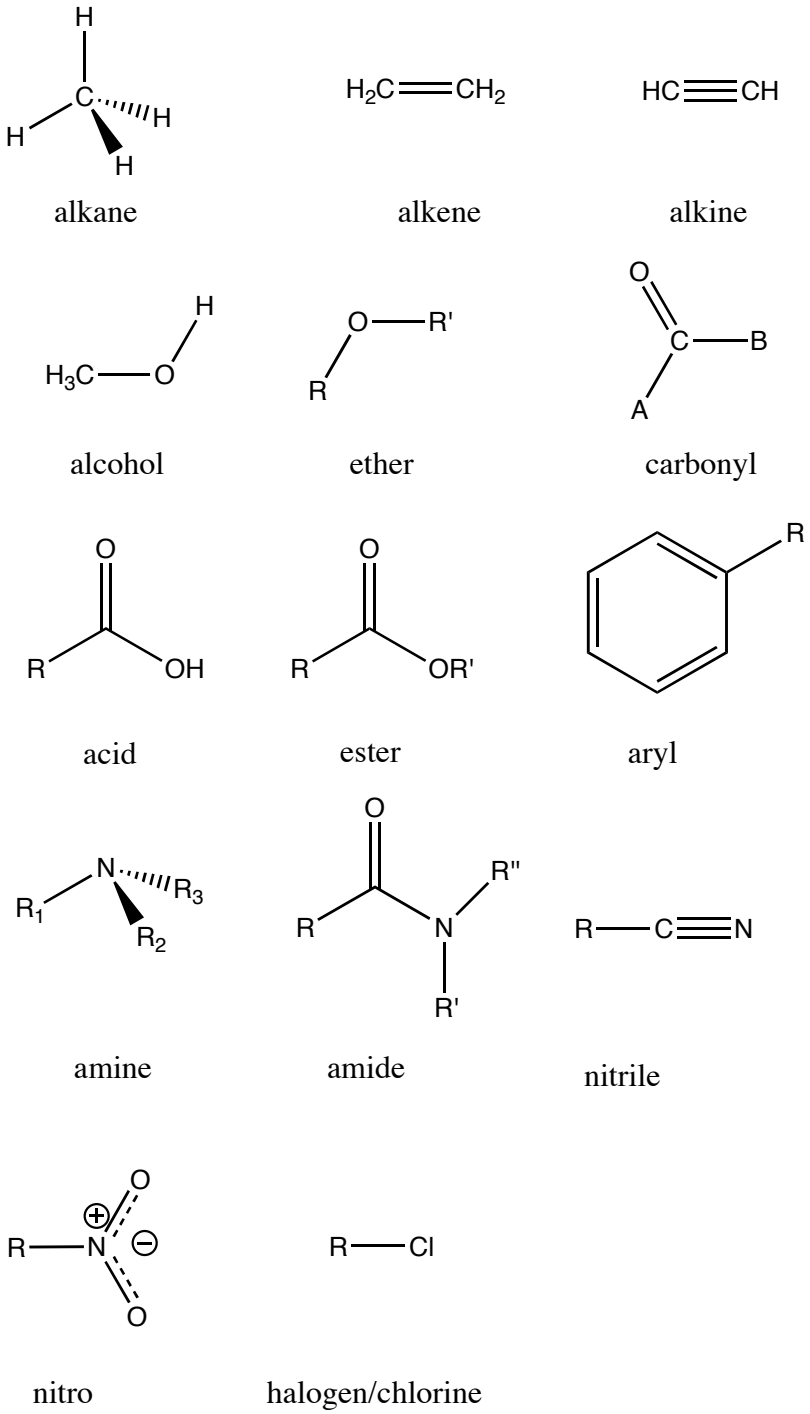


Figure 21: Chemical structure of the different classes of compounds.

Table 12: Average deviation of BP86/TZVP DCOSMO-RS calculated solvation energies compared to experimental hydration energies (kcal/mol) for several functional group categories considered. (there is no experimental results available for the amines class).

group	number of molecules	COSab-DC	COSab-DM	DCOSMO-RS
alkanes	15	-2.92	-2.81	-1.45
alkenes	16	-3.13	-3.07	-1.52
alkynes	8	-4.10	-4.16	-2.81
alcohols	14	-1.87	-1.74	-0.70
ethers	14	-3.37	-3.13	-2.15
carbonyls	18	-2.28	-2.31	-1.36
acids	3	-1.62	-1.53	-1.26
esters	8	-4.73	-4.66	-3.87
aryls	25	-2.25	-2.50	-0.63
amines	20	-	-	-
amides	7	-1.01	-0.90	-0.91
aryls (with N)	16	-1.10	-1.27	-0.25
nitriles	10	-2.79	-2.87	-1.99
nitro	12	-3.22	-3.27	-2.70
halogens	22	-2.93	-2.99	-1.14

One can also break down the data to assess the behavior of the different models for various classes of molecular functionality (Table 12 and Figures 18-20). This is potentially problematic due to the number of experimentally known solvation energies known in each class. Table 12 reports the average difference between the BP86/TZVP DCOSMO-RS calculated solvation energy and the experimental hydration energy, for each class of molecules (see Appendix A for additional details). As such, these values provide a general evaluation of the performance of the solvation models. While many of the classes are within an acceptable range, there are some that are problematic for the model in some way, or, not an adequate number of data to assess in that particular class.

For example, the class of esters has an average error of more than 4 kcal/mol for COSab-DM and COSab-DC methods. Looking at the details (Appendix A), one can see that acetic acid ethyl ester and acetic acid propyl ester, in particular, are more than 5 kcal/mol of error with respect to the experimental results. On the other hand, the DCOSMO-RS method, while still having a quite high error (-3.87 kcal), provides

better results than the other two methods, with the inclusion of non-electrostatic effects.

Another class that has relatively high average error is that of the alkyne functionality, with more than 4 kcal average error, with both COSab-DC and COSab-DM methods. However, again, the DCOSMO-RS method shows less than 3 kcal average error. Looking at the details of the alkyne results, one can see that hex-1-yne and oct-1-yne are both ~5 kcal in error with the two methods that do not include non-electrostatic effects, and only 2.5 kcal in error with the improved, DCOSMO-RS method.

Across all functional group types, the DCOSMO-RS model shows significant improvement over the other two, due to the additional non-electrostatic effects. The extent of the improvement is proportional to the importance of non-electrostatics in that class of systems, as in the above two examples. In general, however, one sees probably a better than expected performance from the COSab-DC and COSab-DM methods, particularly for those functional group classes that are < 2.0 kcal of averaage error (e.g., alcohols, acids, amides, aryls). This is actually quite good for a computational methodology based on a continuum model without the inclusion of nonelectrostatic effects. With the inclusion of non-electrostatic effects, we see many functional group classes that are now within that same 2.0 kcal average error (10 of the 14 groups considered), and only 1 class (esters) has an average error over 3 kcal. Even the alkane class, which is notoriously difficult to calculate using continuum solvent methods, gives an average error of nearly 3 kcal using only the electrostatic solvation methods, but drops to 1.45 kcal average error with the inclusion of non-electrostatic effects, supporting the known importance of such effects in this class of molecules.

Of the 217 molecules in the dataset, 22 deviate from the experimental value by more than 4.0 kcal/mol, using the COSab-DC and COSab-DM model. The outliers include cyclohexane, hex-1-yne, oct-1-yne, but-1-en-3-yne, 1-propoxy-propane, 2-methoxy-2-methyl-propane, dimethoxy-methane, formic acid ethyl ester, acetic acid methyl ester, acetic acid ethyl ester, acetic acid propyl ester, propanoic acid ethyl ester, biphenyl, diphenyl-methane, anthracene, nitromethane, nitro-benzene, 1-methyl-2-nitro-benzene, hexanchloro-ethane, trans-1,2-dichloro-ethene, 1,2-dichloro-benzene. Of the possible reasons for these rather large errors, not all are due to the

theoretical model. In some cases, there are specific solvation phenomenons that make it difficult to predict a good value from theoretical modeling, including the presence of more than one isomeric form, the formation of different molecular forms with water, etc. For example, it is known that formaldehyde forms methanediol in water (Figure 22),[40] which is difficult to take into account in the present model.

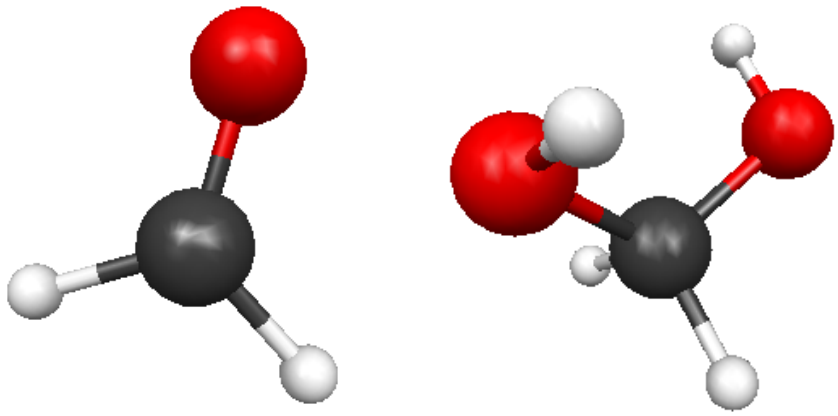


Figure 22: Formaldehyde (left) and Methanediol (right) representation.

3.6.2 Prediction of Acid Dissociation Constant

Acid dissociation constants are of great importance in many areas of organic or biological chemistry. The difficulty associated with computational prediction of such values has motivated a lot a research in the prediction of pKa values in particular.[69-73] Very accurate solvation energies are necessary in order to obtain accurate value of pKa. Extremely small differences in energy (e.g., 1.5 kcal/mol) result in large differences in pK_a, due to the ln dependence (1 pK unit).

The relationship between free energy and pKa is:

$$pK_a = \frac{\Delta G_{diss}}{RT\ln(10)} \qquad \textbf{32}$$

Here, the calculation of free energy of dissociation is defined according to the following relationship:

$$AH + H_2O \rightarrow A^- + H_3O^+$$

where, AH is the acid form of the molecule. Some literature references make use of the full thermodynamical cycle,[69] , as shown in Figure 23. However, the theory presented in this thesis, using DCOSMO-RS allows a more direct analysis, by calculating ΔG_{diss}, directly computing pKa.[72]

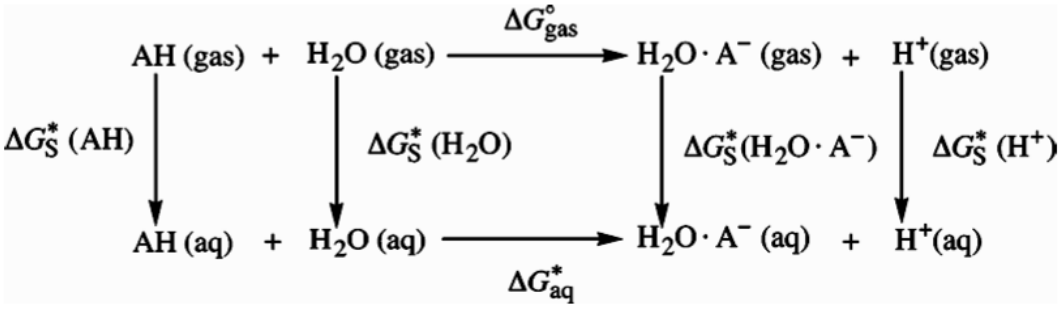


Figure 23: Thermodynamical cycle of acid dissociation.

Klamt et al was first to use the COSMO-RS theory to predict pKa in a direct way,[72,74] We propose here to use the method as implemented in GAMESS, to make predictions of pKa as another test for the COSab methodology. Again, we employ the BP86/TZVP level of theory as before. In addition, the basis set 6-311+G(2d,p) is also used for comparison, as a basis set containing supplementary diffuse functions that may provide better results for anions in particular.

For these computations, the implemented COSab procedure enables the calculation of ΔG of the acid dissociation in solvent. We then use the linear free energy relationship (LFER):

$$pKa = C_1 \Delta Gdiss + C_0 \qquad \textbf{33}$$

where, the 2 factors C_1 and C_0 are the LFER parameters. Using the BP86/TZVP method enables use of already determined parameters for C_1 and C_0 in the post-processing COSMOtherm procedure. These parameters must then be re-determined for other levels of theory.

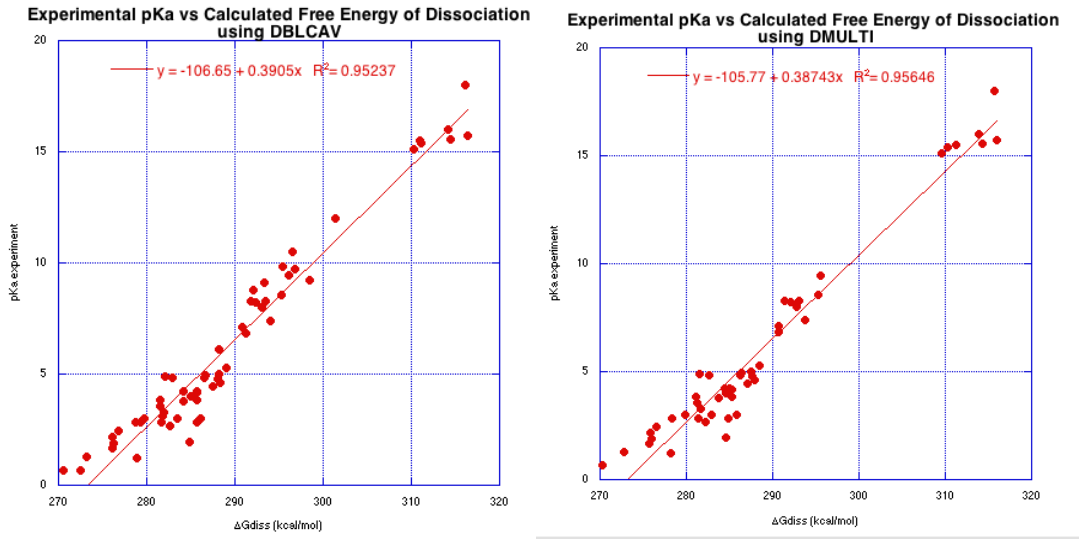


Figure 24: Experimental pKa vs. Calculated free energy of dissociation at BP86/TZVP using double cavity and distributed multipoles.

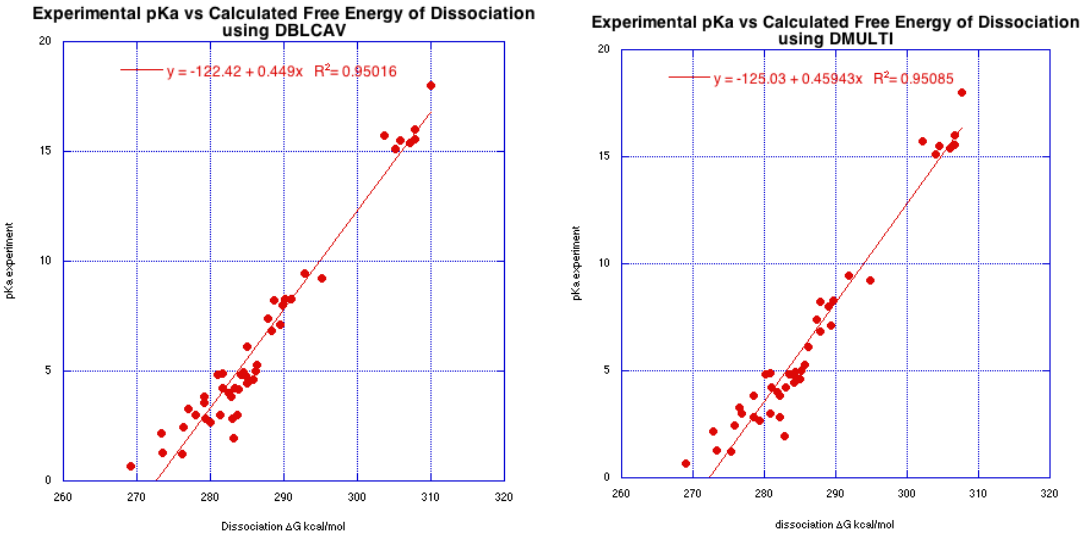


Figure 25: Experimental pKa vs. Calculated free energy of dissociation at BP86/6-311+G(2d,p) using double cavity and distributed multipoles.

The dissociation free energies calculated using the presented method combines the advantages of the quantum mechanical accuracy of the COSab method, as well as thermodynamics treatment of the deviation from the dielectric solvation due to specific local interactions using COSMO-RS (COSMOtherm). Using this combination, the accuracy of the obtain pKa is good even though the theoretical slope of 1/RTln(10) is not respected. This factor constitutes the nonlinearity of the continuum model.

The full set of results, composed of organic molecule used in literature for evaluation of pKa prediction,[71,72] are presented in **Error! Reference source not found.**. In general, it is found that the model does not give satisfactory results for molecules with pKa larger than about 15. This has also been found for other models of this type.[72,74] Using a different method based on a combination of experimental and calculated gas and liquid-phase free energies,[71] Kelly et al. shows significantly decreased accuracy for cases of high pKa values. The two methods, COSab-DM and COSab-DC, both provide similar accuracy of results, using the BP86/TZVP level of theory. This level of theory has also been used for parameterization for the COSMO-RS theory. In this way, COSab is able to produced accurate free energy of solvation (as shown in the previous chapter), but here also shown to provide correct description of the physics of acid dissociation in water.

In the cases where there is more than 1 pKa unit of error in the prediction, one finds that most of these molecules have more than one conformation that should be considered as contributing to the final result. In such cases, one needs to consider all isomers, as well as their contribution to the overall structure, to determine the final result.

Table 13: Acids with Experimental and Calculated pKa values and Calculated Free Energy of Dissociation.

Molecule	ΔGdiss BP86/TZVP dblcav	pKa BP86/TZVP dblcav	ΔGdiss BP86/TZVP dmulti	pKa BP86/TZVP dmulti	ΔGdiss BP86/6-311+G(2d,p) dblcav	ΔGdiss BP86/6-311+G(2d,p) dmulti	pKa expt
acetamide	310.22	15.1	309.60	14.58	305.18	304.08	15.1
aceticacid	288.08	5.11	287.64	4.66	284.87	284.19	4.76
acetyl2butanedioicacid	285.72	4.78	284.81	4.22	283.08	282.16	2.86
acrylicacid	284.11	4.21	284.51	4.22	281.67	280.99	4.25
allyl_alcohol	311.01	15.69	311.23	15.58	305.85	304.49	15.5
anthranilic	286.70	5.87	286.35	5.61	284.47	284.37	4.95
benzoicacid	285.73	5.02	285.04	4.55	283.39	282.93	4.2
benzylalcohol	311.07	16.27	310.35	15.77	307.11	306.09	15.4
boric	298.44	10.06			295.16	294.92	9.23
bromoacetic	278.80	1.88	278.39	1.52			2.86
carbonic	281.60	3.66	281.21	3.35	279.10		3.58
chloro2phenol	291.86	8.83	291.44	8.57	290.24	289.60	8.29
chloro2propanoicacid	281.71	3.4	281.37	3.06	279.31	278.54	2.83
chloro3phenol	292.05	9.32					8.78
chloro3propanoicacid	284.98	4.6	284.60	4.13	282.57	281.81	3.98
chloro4phenol	293.34	9.68					9.14
chloroaceticacid	279.25	2.62					2.86
cinnamic	287.45	5.44	287.04	5.09	285.03	284.21	4.44
cyanoaceticacid	276.82	0.91	276.56	0.69	276.36	275.86	2.43

dichloroacetic	273.23	-0.05	272.81	-0.29	273.59	273.42	1.29
dimethadione	288.16	6.43			285.04	286.10	6.1
dimethyl22propanoicacid	288.18	5.6	287.53	5.11	286.18	285.20	5.03
ethanol	314.18	17.05	313.95	16.81	307.84	306.59	16
fluoroaceticacid	282.61	3.44	282.26	3.18	280.02	279.35	2.66
fluorouracil	293.07	8.48	292.77	8.14	289.89	289.06	8
formicacid	284.13	3.79	283.71				3.77
fumaric	279.78	3.02	279.81	3.01	277.96	276.88	3.02
glycolic	281.52	3.25	281.18	3.03	279.17	278.53	3.84
h2o	316.36	16.67	315.94	16.02	303.71	302.21	15.74
hydroxy2propanoicacid	285.64	4.28	285.24	3.87	282.91	282.15	3.86
hydroxycinnamic	288.31	5.83	287.92	5.5	285.91	285.04	4.61
hypobromousacid	295.30	9.67	295.33	9.55			8.55
hypochlorousacid	294.05	8.65	293.77	8.37	287.86	287.33	7.4
hypoiodousacid	296.58	10.12					10.5
iodoaceticacid	281.81	3.48					3.12
isonicotinic	282.97	4.03	282.65	3.66	280.94	280.21	4.84
maleic	284.90	4.13	284.64	3.91	283.17	282.79	1.93
methanol	314.45	17.16	314.32	16.91	307.87	306.73	15.54
methylthiouracil	292.37	8.31	292.03	7.99	288.60	287.86	8.2
napthoic	285.63	5.04	285.31	4.77	283.81		4.17
nitro5uracil	289.04	6.97	288.52	6.67	286.42	285.60	5.3
nitroaceticacid	276.16	1.69	275.76	1.4			1.68
nitrophenol	290.86	7.54	290.66	7.58	289.59	289.37	7.14
nitrous	282.00	3.62	281.68	3.56	277.05	276.44	3.29
oxalic	278.94	1.78	278.25	1.4	276.11	275.38	1.23
pentachlorophenol	282.04	5.12	281.53	5.4	281.69	280.78	4.9
pentanoicacid	286.49	4.78	286.30	4.49	284.16	280.78	4.84
phenol	295.46	9.98					9.82
phosphoricacid	276.06	2.37	275.78	2.47	273.26	272.89	2.16
phtalamide	293.52	8.83	293.02	8.57	291.08	289.71	8.3
phthalicacid	283.44	3.92	282.94	3.63	281.39	280.76	2.98
salicylicacid	286.18	5.28	285.81	5.01	283.67		2.98
sulfurousacid	276.19	1.9	275.97	1.76			1.9
tbutylalcohol	316.18	18.21	315.72	17.82	310.07	307.74	18
thymine	296.80	9.76					9.75
trans5formyluracil	291.29	7.69	290.74	7.34	288.30	287.84	6.84
trichloro222ethanol	301.38	12.61					12.02
trichloroacetic	270.57	0.98	270.23	-0.72	269.17	268.93	0.65
uracil	296.07	9.43	295.52	8.9	292.91	291.84	9.42

Solvents other than water can also be explored. Figure 26 shows calculated results for dimethyl sulfoxide (DMSO), with ε=48. For these computations, as explained earlier, the LFER parameters are different from those used with water solvent. Using the BP86/TZVP level of theory, one obtains a standard set of parameters, which are used in the COSMOtherm software. The resulting correlation is found to be good, R^2>0.97, as also previously found by Klamt.[72]

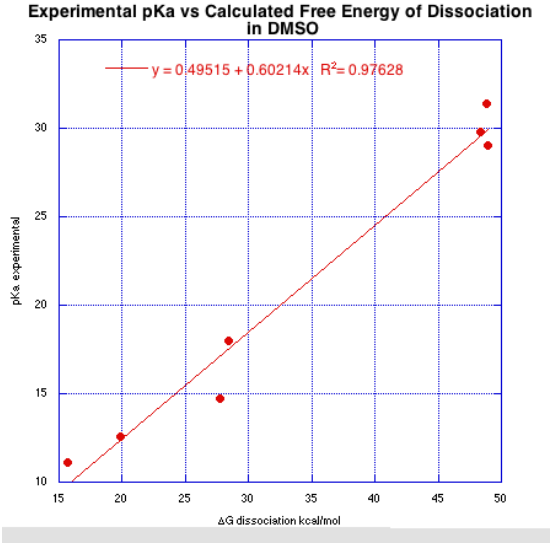


Figure 26: Experimental pKa vs. Calculated free energy of dissociation in DMSO at BP86/TZVP using double cavity.

We conclude that the implemented method using the COSab/COSMO-RS combined theory is a convenient method to predict acid dissociation constant in infinite dilution of an acid in pure water, as well as in other solvent. The main advantages include the avoidance of known computational problems associated with the solution phase computations within the thermodynamics cycle, as well as significant savings in computational time.

3.6.3 Application of Sigma Profiles

3.6.3.1 Benzene Isomers

Faraday first discovered benzene in 1825, however its structure was determined much later, by Kekule in 1865 years.[75] Theoretically, there are 330 isomers with the molecular formula C_6H_6, of which 217 can be excluded as diastereomers and/or enantiomers.[76]

Five valence isomers of C_6H_6, in particular, attract attention as the primary isomeric forms of benzene: benzevalene, dewar benzene, prismane, and bicyclopropenyl. Dewar benzene was synthesized in 1963 by van Tamelen and Pappas.[77] Wilzbach et al. isolated benzvalene in 1967,[78] and Katz et al. synthesized the same molecule in 1971.[79] Prismane was prepared from benzvalene by Katz and Acton in 1973.[80] Derivatives of bicycloprop-2-enyl have been prepared by Breslow and Davis et al., while the synthesis of the parent isomer was achieved in 1989 by Billups and Haley.[81] These isomers of benzene are illustrated in Figure 27.

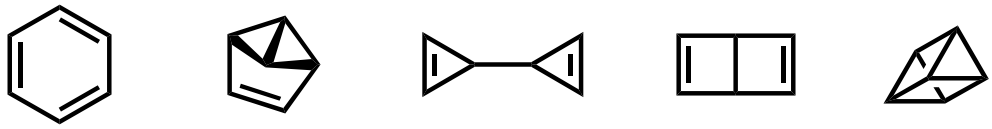


Figure 27. C_6H_6 isomers. From left to right: benzene, benzvalene, bicyclopropenyl, dewar benzene and prismane.

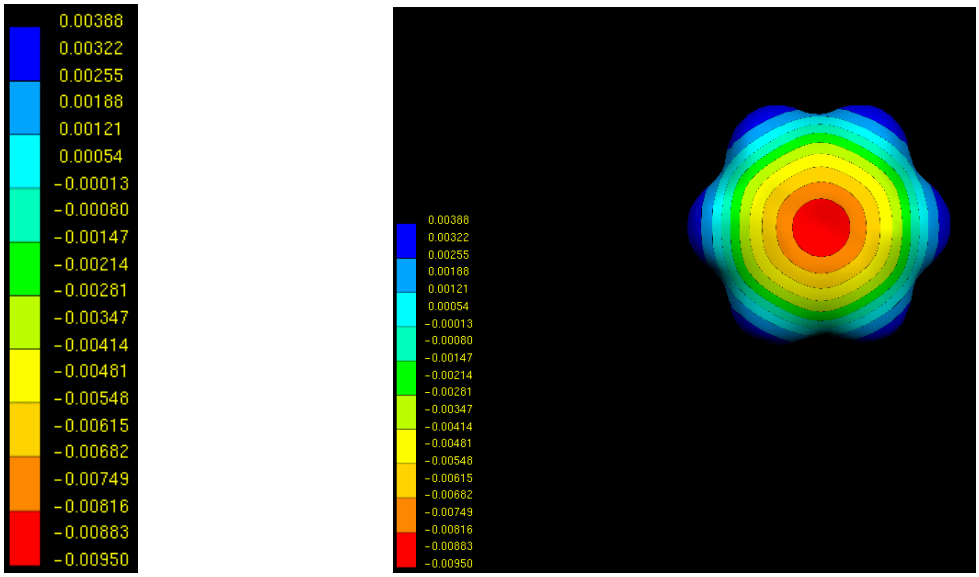
While the structural variance in these five isomers is clear from their geometric construction, their electronic structure and reactivity is less obvious. We can, however, investigate the five isomers using the theories presented in this work, in particular their σ-profiles. In gas phase, a basic study of their properties has been done, using the B3LYP/cc-pVDZ and MP2/cc-pVDZ levels of theory. These results are presented in Table 13. Despite their similar molecular weight, the presented isomers have very different dipoles, which mean their behavior in liquid phase would be very different. Additionally, the predicted HOMO (eV) values, which relate to their expected ionization potentials (as per Koopmans Theorem[82] estimation), also show their quite variant reactivity.

Table 14: Energetics and Properties of C6H6 isomers using B3LYP/cc-pVDZ and MP2/cc-pVDZ optimization. Relative Energy in kcal/mol.

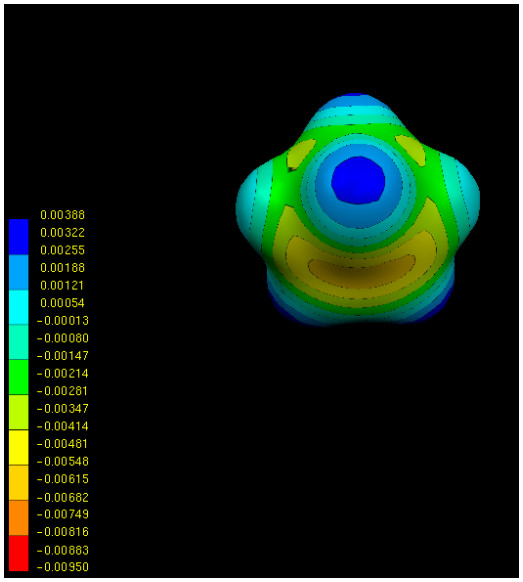
B3LYP opt	**benzene**	**benzvalene**	**dewar benzene**	**bicyclopropenyl**	**prismane**
HOMO (eV)	-6.80	-5.82	-6.56	-6.46	-6.14
LUMO (eV)	1.62	0.37	0.99	0.97	1.96
Dipole (D)	0.00	0.85	0.01	0.72	0.00
rel. E.	0.00	81.49	84.71	130.26	121.91
MP2 opt					
HOMO eV	-9.03	-8.49	-9.09	-9.26	-9.22
LUMO eV	3.67	4.39	3.69	3.80	5.26
Dipole (D)	0.00	0.89	0.06	0.71	0.00
rel. E.	0.00	74.92	81.10	129.40	116.83

Another way to understand the variation in reactivity of these isomers, involves comparison of their respective molecular electrostatic potential maps, providing a visual method to understand the response of the molecule to an external charge. The molecular electrostatic potential is the response of a molecule to a proton at a particular location near a molecule: negative corresponds to an attraction of the proton (red), and positive corresponds to a repulsion of the proton (blue). Figure 28 shows the results of molecular electrostatic potential (MEP) maps for the five isomers. The MEP maps are presented using the same scale for comparison purpose. The scale is represented below.

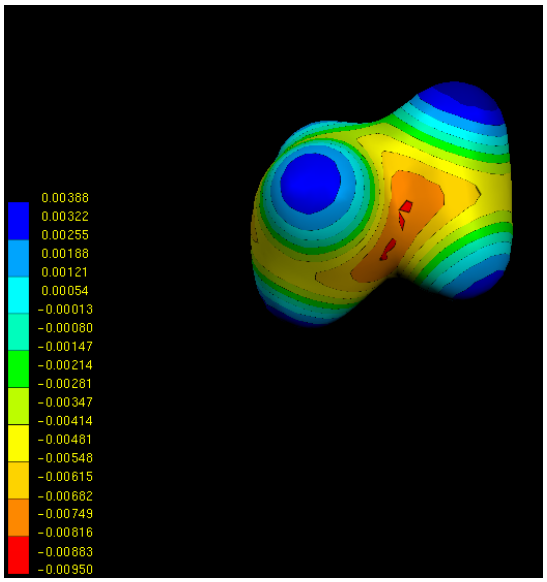
a)



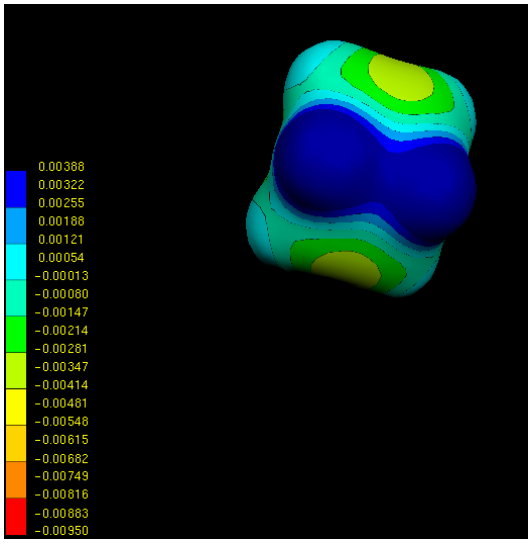
b)


0.00388
0.00322
0.00255
0.00188
0.00121
0.00054
-0.00013
-0.00080
-0.00147
-0.00214
-0.00281
-0.00347
-0.00414
-0.00481
-0.00548
-0.00615
-0.00682
-0.00749
-0.00816
-0.00883
-0.00950

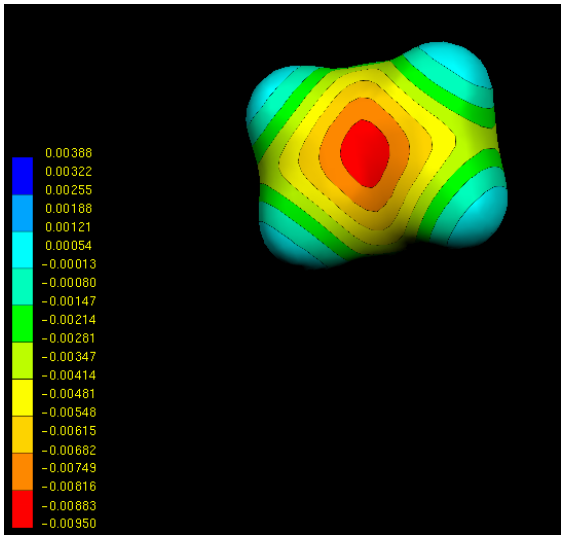
c)


0.00388
0.00322
0.00255
0.00188
0.00121
0.00054
-0.00013
-0.00080
-0.00147
-0.00214
-0.00281
-0.00347
-0.00414
-0.00481
-0.00548
-0.00615
-0.00682
-0.00749
-0.00816
-0.00883
-0.00950

d)


0.00388
0.00322
0.00255
0.00188
0.00121
0.00054
-0.00013
-0.00080
-0.00147
-0.00214
-0.00281
-0.00347
-0.00414
-0.00481
-0.00548
-0.00615
-0.00682
-0.00749
-0.00816
-0.00883
-0.00950

e)


0.00388
0.00322
0.00255
0.00188
0.00121
0.00054
-0.00013
-0.00080
-0.00147
-0.00214
-0.00281
-0.00347
-0.00414
-0.00481
-0.00548
-0.00615
-0.00682
-0.00749
-0.00816
-0.00883
-0.00950

f) g)

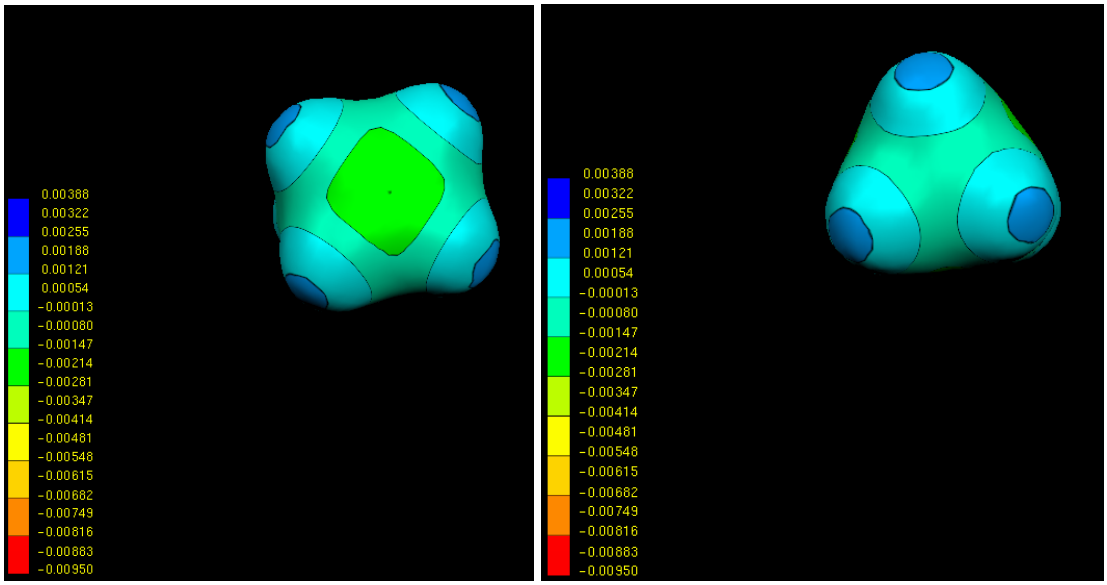


Figure 28: Molecular Electrostatic Potential map of C_6H_6 isomers from MP2/cc-pVDZ geometry optimization calculations. a) benzene b) benzvalene c) bicyclopronenyl d) Dewar benzene from top e) Dewar benzene from bottom f) prismane from the bottom g) prismane from the side. The scale is the same across the set.

From the molecular electrostatic map, it is directly visible that the electronic distribution is very different across the set of isomers of benzene. For example, for benzene or Dewar benzene, there are significant electronegative areas, compared to prismane on the same scale showing no highly electronegative areas. The difference in structure manifests in highly different electronic distribution patterns, and therefore very different reactivity patterns.

Analogously, one can determine the expected behavior of these isomers in a solvent environment. We have used the theory presented previously to calculate solvation properties for the set of C_6H_6 isomers, as summarized in Table 14. Inspection of their gas phase dipole moment provides insight into their expected response in solution environment. However, Table 14 shows that all of the isomers have dipoles less than 1 D, with benzvalene and bicyclopropenyl having the largest values (0.8 D and 0.7 D). The solvation energy, as predicted by the COSab theory, are in fact, very similar in the five isomers, with benzvalene having the largest value. However, it is not the case that bicyclopropenyl is the second largest, as might be

predicted by the dipole effect alone. One also must consider the available surface area for interaction with a particular solvent.

The data provided in Table 15 also shows both the volume as well as surface area of the cavity for the different isomers. The volume of the bicyclopropenyl is quit a bit larger than the other isomers, and prismane the smallest isomer by this measure. Benzene, benzvalene and Dewar benzene are all about the same volume. The surface area follows this same pattern. In order to understand more detailed information regarding their expected reactivity patterns with respect to solvent, we appeal to the sigma profiles, as presented earlier.

Table 15: Solvation properties for C_6H_6 isomers.

BP86/TZVP COSab-DC	**benzene**	**benzvalene**	**dewar benzene**	**bicyclopropenyl**	**prismane**
Rel. E. corrected (kcal/mol)	0.00	76.57	81.91	123.92	117.90
ΔE_{elect} solvation (kcal/mol)	-3.08	-3.37	-2.20	-2.64	-2.53
Outlying charge effect (kcal/mol)	-0.0001	-0.0001	0.0000	0.0000	0.0001
Dielectric Energy[a.u.]	-0.0055	-0.0059	-0.0041	-0.0054	-0.0039
Sum of polarization charges	0.0275	0.0264	0.0267	0.0253	0.0253
Surface area of cavity (A**2)	433.13	421.72	433.48	460.79	415.23
Volume of cavity (A**3)	741.61	725.19	746.51	800.48	713.89
Number of segments	522	451	500	520	465

Figure 29 shows the sigma profiles generated from the solvent calculations for each of the isomers.

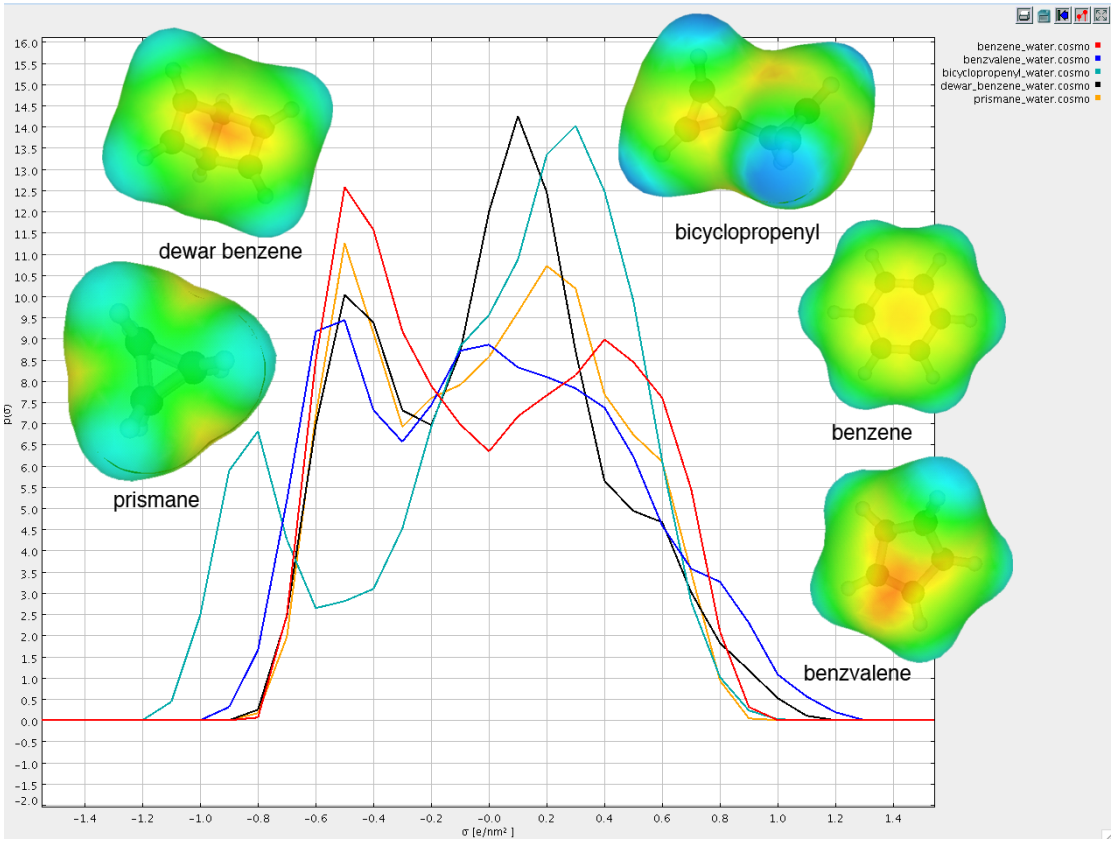


Figure 29: sigma-profiles of selected c6h6 isomers: benzene (red), benzvalene (dark blue), bicyclopropenyl (green), Dewar benzene (black), prismane (yellow).

For all cases, the two primary peaks are on either side of the central, non-polar part of the graph. However, the actual position of the peak, the height of the peak, and the broadening of these two peaks all highlight the differences b), and the partial negative polarity of the pi-face of the benzene ring between the various isomers. These subtle differences in electronic distribution of the molecules make them react differently in the aqueous phase. The peaks distributions provide a description of the polarity of the molecules surface, making it possible to observe even small differences in solute-solvent interaction potential. The σ-profiles for the set of isomers in general show two peaks each, arising from the exposed carbon and hydrogen surfaces, respectively. The actual position of the peaks varies with the electronic structure of the particular conformation. Looking first at benzene, the σ-profile shows two well-separated peaks on either side of the central, nonpolar region of the graph. The peak on the negative (left) side corresponds to the partial positive polarity of the hydrogen belt of benzene, (-0.5 e/nm^2), and the peak on the positive (right) side corresponds to the partial negative charge from the π-face of the benzene ring (+0.4 e/nm^2). The

height of the peaks corresponds to the size of the surface that has a particular σ -profile value, and the width of the peak shows the amount of the surface that has gradual variation from the top of the peak. For example, there is less area associated with the partial negative π-region surface of benzene than the partial positive belt region. The π-face corresponds to the quadrupolar region at the center of the ring and is relatively broad around the peak region. On the other hand, the partial positive belt of benzene is a much larger region, and has a correspondingly larger but narrow peak. ^{2}). This analogue and the Dewar benzene analog have the largest amount of area of partial negative charge (+0.3 and +0.05 e/nm^2), although smaller magnitude than benzene. The benzvalene isomer has the most attenuated and broad partial negative charge surface. These differences in surface ultimately correspond to how the molecule will interact with other species. For example, more focused negative charge will have more focused interaction with positive components of interacting molecules. A simple dielectric calculation with the single measure of solvation energy is not able to differentiate between isomers or predict interaction potential of a molecule. The subtle difference in electron distribution among different isomers is difficult to determine using standard methods, as the differences in energy are minimal. Using diagnostic tools like the σ-profile, such differences become clearer.

3.6.3.2 Halide Series

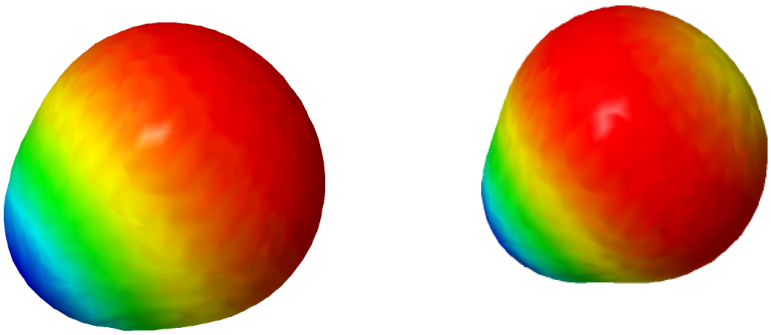
In the previous chapter, we used the σ-profiles to detect differences among isomers, which have small variations in molecular structure. We now demonstrate the use of the sigma profiles to observe electronegativity differences within a set of molecules. This set is composed of hydrogen fluoride (HF), hydrogen chloride (HCl), hydrobromic acid (HBr) and hydrogen iodide (HI). We again use the standard BP86/TZVP method to do gas phase optimizations, as well as optimization in water environment, ε=78.4. Table 16 summarizes the results.

Table 16: Solvation Energy and Property of Halide Series. ΔE_{elect} and OCE are in kcal/mol, Dielectric Energy in a.u., volume of the cavity in A3, and dipole in Debye.

Molecule	ΔE_{elect} so	OCE	Dielectric Energy	cavity volume	dipole gas phase	dipole in solvent
HF	-6.05	-0.0734	-0.0106	156.98	1.94	2.22
HCl	-3.69	-0.1166	-0.0067	255.95	1.41	1.80
HBr	-2.87	-0.1286	-0.0054	297.94	1.06	1.46
HI	-2.35	-0.1116	-0.0043	366.61	0.73	1.11

Initially, gas phase molecular electrostatic potential maps have been generated for each molecule, as shown in Figure 30, to highlight the differences in electron density distribution in these molecules. As is well known, fluorine is the most electronegative of the set, with chlorine, bromine and iodine progressively more electropositive. The interesting observation is at the very tip of the halogen surface, where one sees a trend towards larger and larger positive contour with increasing size of the halogen species. This has been observed in the literature[83] and postulated as contributing to the possibility the larger halogen atoms participating in what are termed 'halogen bonds', similar to hydrogen bonds. The phenomenon has been particularly documented in biological system. [83-85]

a) b)



c) d)

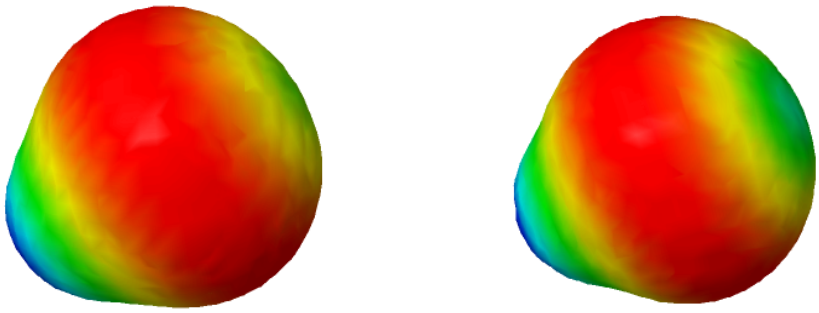


Figure 30: Molecular electrostatic potential map of the halide series. a) HF b) HCl c) HBr d) HI.

As before, for more detail concerning the difference in reactivity of these molecules in a solvent environment, we can carry out solvent calculations and look at their respective σ-profiles, as shown in Figure 31. For example, comparing the sigma surfaces of HF through HI, one observes a difference in the width of the sigma profile curve, decreasing with decrease in electronegativity of the halogen. Additionally, the apolar component increases in intensity, as also the gas phase MEP previously showed. The σ-surface shown in the small 3D plots, illustrates this difference in polarity. HF (right side) is very polar, as compared to HI (left side) is less polar.

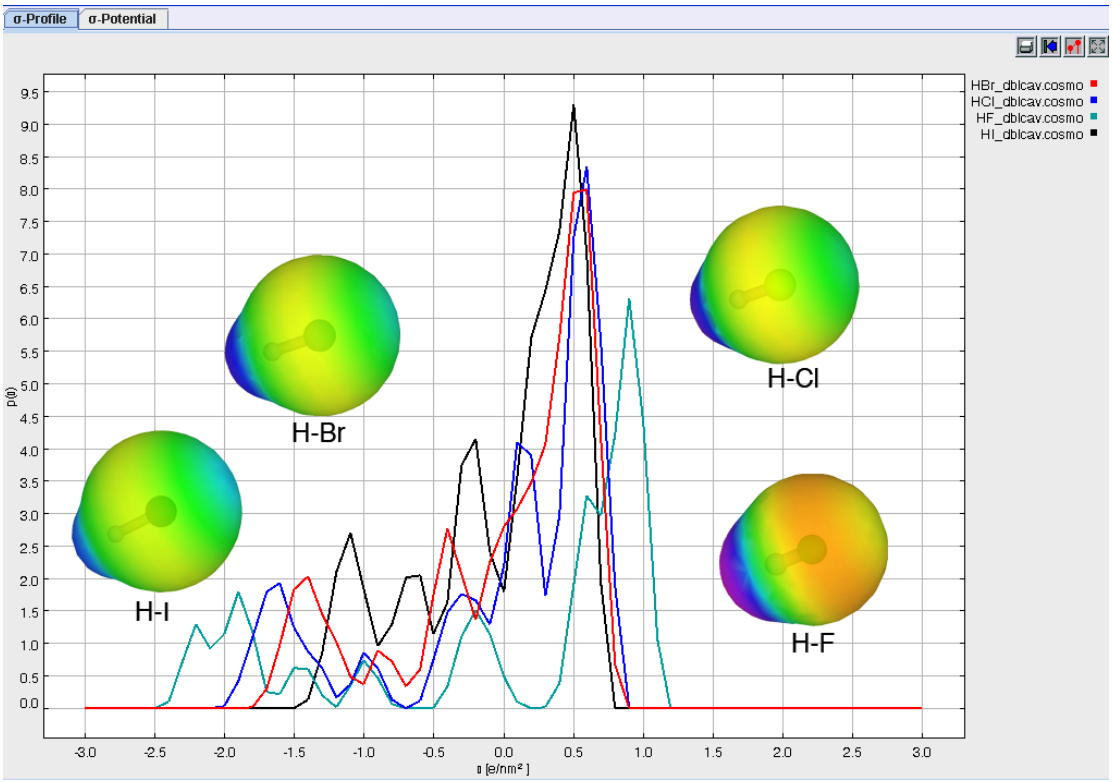


Figure 31: Sigma-profiles of halides series. HF: light green, HCl: dark blue, HBr: red, HI: black. The σ-surface of HF (right) and HI (left) are also displayed.

The threshold for a hydrogen bond is normal defined in this theory at σ= ±0.79e/nm^2. However, hydrogen bonding is generally considered as 'weak' up to ±1e/nm^2. On the positive side of the sigma-profile plot, only hydrogen fluoride is beyond this limit. This also is in accord with the definition of hydrogen bond given by Pauling[86]: the strength of hydrogen-halogen bonds should increase with increase in the electronegativity of the halogen. In general, we expect fluorine to be the more electronegative of the halogen family, decreasing down the column of halogens. The increasing size down the halogen family results in an increase H-X distance, and a decreasing hydrogen bond energy. Finally, we observe on the sigma surface a progressive change in polarity along the bond, emphasizing how important the description of the cavity is to obtain an accurate description of the electronic structure, and therefore response properties, around the cavity.

In conclusion, we find that the computational method can benefit in better understanding and differentiation among similar species with different electronic properties. The difference in polarizability among molecules can be observed in more detail with such tools.

3.7 Conclusion

In the presented methodology and associated analysis tools, the chemical treatments of molecular and electronic structure with inclusion of solvent effects directly in the *Fock* matrix enables prediction of geometry as well as electronic properties, with a high degree of detail.

In the present chapter, we show comparison and implication for capturing the outlying charge effect with two different methods. Both strategies prove to be accurate method for correcting this error. Applications of both models have been presented with a variety of quantum mechanical methods. The cavity construction and surface charge evaluation have been improved over that of previous implementations. The physics of the model has been tested and verified. The model has been applied to the prediction of free energies of solvation for a wide variety of test systems, and shown to be reliable in the most general context. Consideration of non-electrostatic effects has been evaluated using additional tools, in particular with the COSMOtherm software as a post- analysis tool,, resulting in the DCOSMO-RS method. Moreover, a self-consistent implementation of nonelectrostatic effects could be implemented into GAMESS, D-COSab, the intermediary use of the COSMOtherm tool. Several case studies, such as acidity constant prediction, isomeric distinction, and distinctions in a series of halogen halides, were used to illustrate the performance of the developed tools. The use of σ-profile as an aid in determining chemical signatures of molecule is seen to be a very useful predictive tool.

4 Crystal Packing Environments

4.1 Introduction

The forces that hold molecules together in solutions and solids derive from the individual molecules that make up the aggregate.[87,88] As discussed in detail in Chapter 2, there are various forces that are responsible for giving rise to intermolecular cohesion.[89] A realistic description of the intermolecular potential for molecules arranged in a crystal structure represents a challenging task. In order to evaluate properties in crystalline environments, both classical and quantum mechanical theories have been applied. [29,30,38,90] In this work, it was desirable to develop a procedure based on a hybrid QM technique.

OpiX is a program package for the calculation of molecular and intermolecular properties of organic molecules and crystals; it was developed by A. Gavezzotti from the University of Milan. The theory has been also used for other types of systems.[91-94] OPiX consists of several modules that make use of flat input files describing the molecular and crystal information for the systems of interest. The scheme of the different modules with the corresponding input files[95-97] is shown in Figure 32.

A possibility to provide the localization of intermolecular bonding using fully delocalized electron-density cloud is given with the Semi-Classical Density Sums (SCDS-Pixel) approach implemented in OpiX. The calculation of intermolecular energies starts by obtaining the electron density for the separate molecules. Using standard quantum mechanical calculations, a grid usually consisting of points (pixel) is build. This grid is condensed into *n* x *n* x *n* super-pixels, where n is usually 3,4 or 5. Super-pixels with less than 10^{-6} electrons are discarded and the electronic charge is renormalized to balance the total nuclear charge. Finally, the molecular electron density is described by some 10000 to 15000 pixels. The position of all pixels and all nuclei are repeated in space by rotation and/or translation operations of the space group of interest. The coulombic energy between two molecules is calculated simply with the contributions from each pair of electron-density pixels, or in the pixels and nuclei in the separate molecules. The repulsion energy is taken to be proportional to the overlap integral of the electron densities. For the polarization energy, each

electron-density pixel is allocated to an atom depending on the distance and the pixel polarizability is taken as the corresponding atom polarizability, scaled by the ratio of the pixel charge to the atomic number. The polarization energy is then calculated as a many-body effect using the induced dipole over all pixels. The dispersion energy is obtained as a sum of pixel-pixel terms by a London-type formula by using pixels polarizabilities and the overall molecular ionization potential. Polarization and dispersion energies are multiplied by an appropriate damping function to avoid singularities. The details of the method is describe with equation in a section later.[29,30,38,90]

One can prepare an input file from OpiX by hand, or using an existing crystal structure in the format of a Crystallographic Information File (CIF file), as taken from the Cambridge Structural Database (CSD). Figure 32 shows a full description of all the OPiX modules.

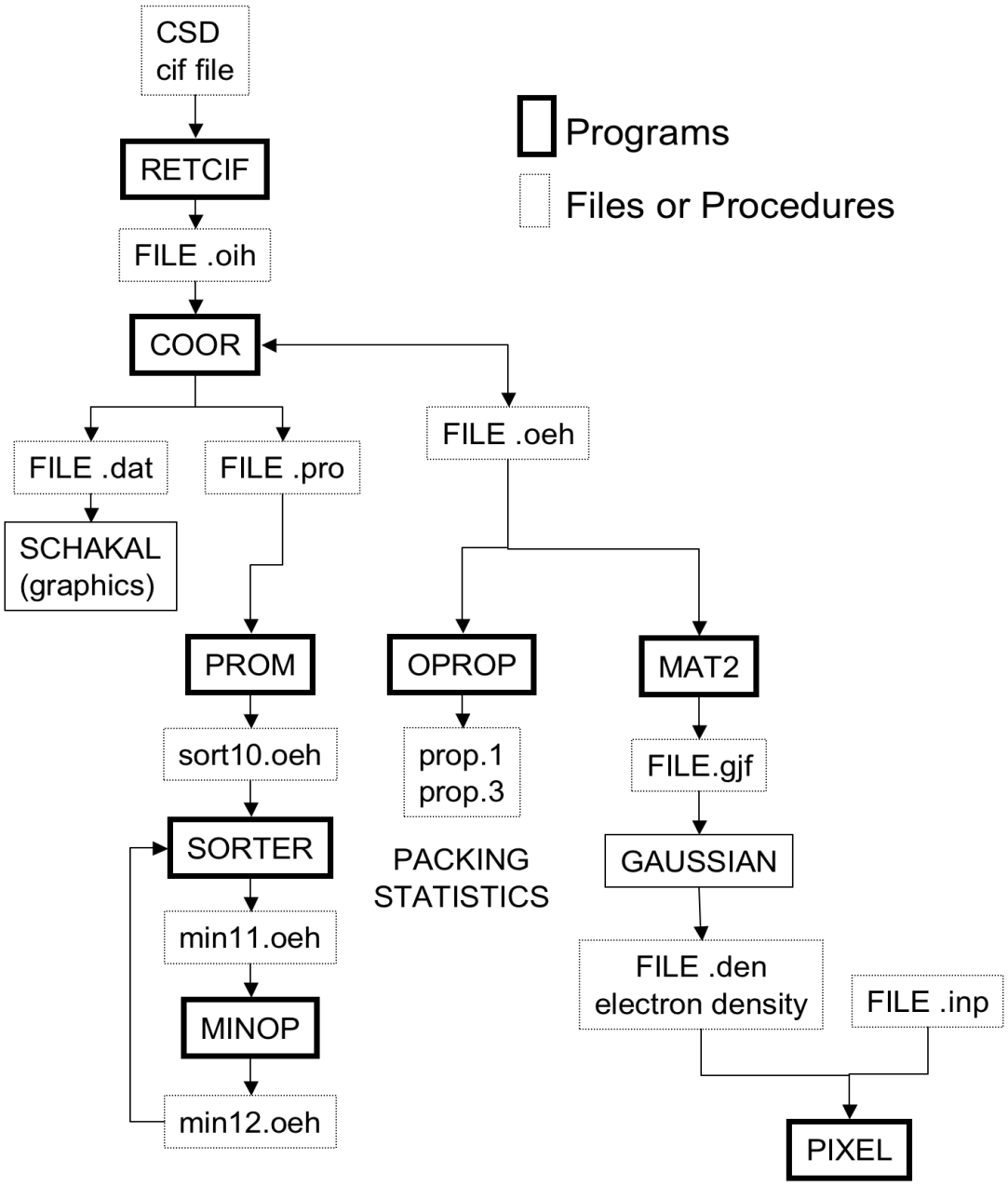


Figure 32: General Overview of the OPiX program package.

4.2 Description of the OPiX Modules

4.2.1 RETCIF Module

The RETCIF module is only a tool to retrieve crystal structure from CIF file from the CSD. The CIF input file provides atomic descriptor, crystal packing

information and cell data for the system of interest. It can also contain information for more than one molecule, or only a fraction of it, depending on the starting crystal structure.

In general, it is possible to have different format for input files: (i) by explicitly assigning hydrogen atoms positions (oeh files) or (ii) by providing a set of possible configurations (oih file).

4.2.2 COOR Module

The COOR module is used to prepare input files for other programs. It makes use of the oih files generated by RETCIF (or made by hand), and provides files for the other modules composing OPiX. It checks for subgroups of the crystal space group and corrects potential errors in the oih file. For example, if the molecule is centrosymmetric and the molecular center of symmetry coincides with a crystallographic center of symmetry, the centrosymmetric space group operations will be deleted in the output file (oeh file).

The visual program SCHAKAL can read the dat file generated by COOR. SCHAKAL is a visual program developed at the University of Frieburg for representation of molecular and solid-state structure models.

The COOR module provides also the input file necessary for the PROM module execution.

4.2.3 PROM Module

Prom is a polymorph generator; it provides a number of crystal structures with the corresponding lattice energies using empirical force fields.

Prom reads a molecular model from the pro file generated from COOR and generates crystal structure for that molecule by using space group operations. For each of the generated structures, the lattice energy is calculated and a series of possible polymorphs is provided. The methods are purely geometrical and the potentials are empirical. The molecule is considered as rigid entity and intramolecular energies are not considered. The only available space groups are the following: P1-, $P2_1$, $P2_12_12_1$, $P2_1/c$, C2/c and Pbca (using Hermann-Mauguin notation). To run Prom, only one fragment must be present in the asymmetric unit. If the fragment consists of more than one molecule (like for salts, solvate crystals or molecular complexes) the

positions of all the different molecules remain fixed and no molecule-molecule interaction energy is calculated inside the unit cell.

Prom is does not enable apriori prediction of crystal structures. In some cases, the minimum energy structure does not correspond to that observed in experiments. The packing arrangement can be observed in the generated crystalline structure, like parallel arrangement of chains or stacking of aromatic rings.[98,99]

Prom enables the building of the three-dimensional crystal structure, according to the assigned symmetry operations. Since a lot of duplicate structures can be generated, the output structures must be clustered and sorted using Sorter. Subsequently, an optimization with respect to rigid-body molecular degrees of freedom and cell parameters can be performed using Minop.

Summarizing: i) Prom generates crystal structures, ii) Sorter detects and eliminate duplicates structures, iii) Minop optimizes lattices energies with respect to cell parameters and rigid-body degrees of freedom.

It is worth noting that space groups such as P1- or $P2_1$ are less demanding computationally to simulate because they involve only one symmetry operation, while space group such as Pbca or C2/c result from a combination of three symmetry operators.

4.2.4 Oprop Module

The Oprop module reads an oeh file. This file provides the coordinates of the reference molecular group (RMG) consisting of a single molecule or several molecules. In the case of a crystal structure, a model of the crystal (SMG: surrounding molecular groups) is built according to the symmetry operations of the space group. Intermolecular distances, crystal densities or packing coefficients are properly checked with this tool.

The Oprop module provides the following intramolecular information:

Intramolecular bond distances and angles

Molecular volume

Molecular surface area

Main moment of inertia

Atomic point charges using Mulliken population analysis on an extended Hückel calculation (when charges are not provided in the oeh file).

Oprop does not calculate intramolecular energies. If there is more than one fragment in the RMG, Oprop calculates the interaction energies between fragments pairs.

When there is only one fragment in the RMG, the total intermolecular non-bonded potential energy of the molecule in the crystal cluster is calculated as:

$$E(pot,pot) = \sum_i \sum_j E(i,j) \qquad \textbf{34}$$

where E(i,j) is represented by a revised '6-exp' potential[100]:

$$E(i,j) = A\exp(-BR_{ij}) - CR_{ij}^{(-6)} + {}^{q_i q_j}\!/_{R_{ij}} \qquad \textbf{35}$$

$$E(latt) = -\Delta H(subl) = {}^{1}\!/_{2}\, E \qquad \textbf{36}$$

where i labels any atom in the RMG and j label any atom in any SMG. E is the potential energy of one mole of molecule in the crystal, while 1/2E represents the gain in energy when one mole of molecules from infinity is brought into contact with the crystal (sublimation energy here).

If there is more than one fragment in the asymmetric unit, Oprop provides the following intermolecular information for each atom of the RMG:

1) All the short intermolecular atom-atom distances

2) The part of the total packing energy due to that fragment (Equation 35 with i and j belonging to RMG)

And for each nearest neighbor molecular pair in the crystal:

1) Distance from center of mass of each fragment in the SMG to center of mass of each fragment in the RMG

2) Molecule-molecule or fragment interaction energy

3) The corresponding angle between axes of maximum inertia

4) The molecular and cell dipoles are calculated and the correction to the lattice energy due to the net cell dipole is estimated.

4.2.5 Pixel Module

The electron density $\rho(x,y,z)$ for a molecule can be calculated on a three dimensional grid with spacing d, with (x,y,z) the coordinates of the center of a small element of volume d^3. This small element of electron density is called "pixel," in analogy to the two dimensional case. The charge within the cubic volume of the pixel i is given by $q_i = \rho_i(x,y,z)d^3$.

The Pixel modules (pixelc and pixeld) calculate coulombic, polarization, dispersion and repulsion energies between separate rigid molecules. Intramolecular energies are not calculated. The Pixel module uses molecular electron densities calculated from quantum mechanics, the current version capable of only using the computational package GAUSSIAN[101] at this time. Each elementary density volume with the corresponding charge is called a pixel.

The Pixel module is made up of two different subprograms:

i) Pixeld - for clusters made of any number of molecules of up to two molecular species types: a solute and a solvent. The position of each molecule in the cluster is obtained transforming the original coordinates for nuclei and pixels by means of an orientation matrix, constructed from three Euler angles and a displacement vector.

ii) Pixelc - for crystal made of up to two molecular species per asymmetric unit. The matrix/vector operation that transforms from the molecular reference frame to coordinates in the crystal structure is obtained by use of the Mat2 module. Starting from the positions of molecules in the fragment Mat2 generates automatically the crystal structure using cell parameters, space group matrices, and a cutoff value for the distance between centers of mass of the central and surrounding molecules.

Our main interest in the current work is with the module pixelc for crystal structure. This module offers the possibility of optimizing lattice energies of crystals with one fragment (up to a maximum of two molecules) in the asymmetric unit.

Pixelc provides the following energy values for each fragment: electrostatic, polarization, repulsion, dispersion and total. The output also provides the 6-exp force field energies (Equation 37, a simpler form of Equation 35): repulsion, attraction, and

total point-charge energy. And in the case of crystals, a factor of 1/2 is applied so that these numbers are the computational equivalent of the enthalpy of sublimation.

$$E_{ij} = A\exp(-BR_{ij}) - CR_{ij}^{-6} \qquad \textbf{37}$$

The pixel output can be used to evaluate the different contributions to the total energy, and to identify the most important interactions between molecules arranged in the crystalline environment. The basic methodology of the Pixel method, referred to as semi-classical density sums (SCDS), has been widely applied to compare lattice energies and sublimation enthalpies.[29,90] We present in what follows the application of the OPiX package to the study structure and energetics of three well known organic molecules in crystal environments.

4.3 Illustrated Examples

A series of three well-studied simple organic molecules in crystalline environments: benzoquinone, difluorobenzene and dichlorobenzene (Figure 33). These have been selected to illustrate the capabilities of the OPiX package.

1,4-Benzoquinone (IUPAC: Cyclohexa-2,5-diene-1,4-dione) of chemical formula CH4O2. The crystal structure for this compound was determined in 1935 by Robertson,[102] and then later refined by Trotter.[103] The structure (CIF file) has been taken from a work of Bolte et al.[104] The favored space group of 1,4-benzoquinone, according to the latest reference, is $P2_1/c$. The corresponding representation of the crystalline packing in accord with the described software is illustrated in Figure 34.

The crystal structure for 1,4-difluorobenzene, $C_6H_4F_2$, was taken from the work of Thalladi et al,[85] and corresponds to monoclinic $P2_1/c$ space group. The corresponding representation of the crystalline packing, as determined from the described software, is shown in Figure 35.

1,4-dichlorobenzene, $C_6H_4Cl_2$, has three polymorphs depending on the temperature: the P1 form crystallizes from the melt at 328 K. At 304 K, the structure transforms to the $P2_1/a$ form, and below 203 K, the structure crystallizes in $P2_1/c$. The structure with triclinic P 21/a space group was selected for illustration, since it is the most common form at room temperature.[4] The representation of the crystalline packing of the molecule as calculated with the hybrid method, is shown in Figure 36.

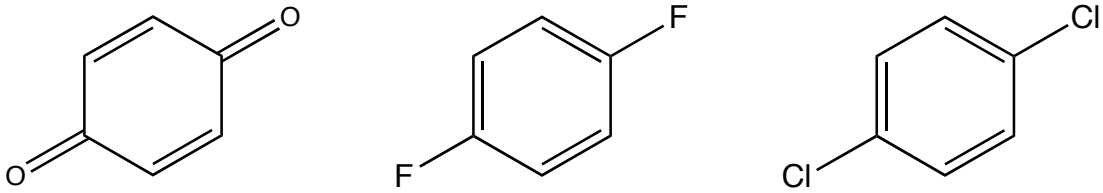


Figure 33: Structure of p-benzoquinone, p-difluorobenzene and p-dichlorobenzene.

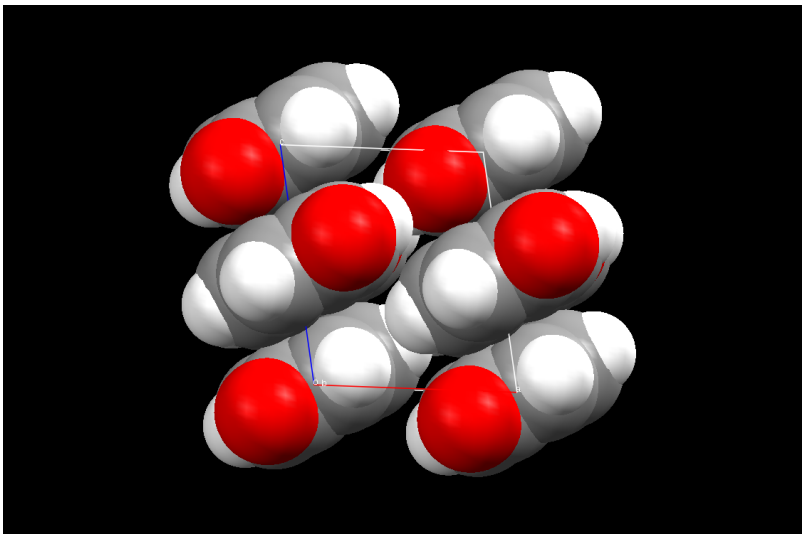


Figure 34: Crystal structure of 1,4-benzoquinone, $P2_1/c$ Z=2.

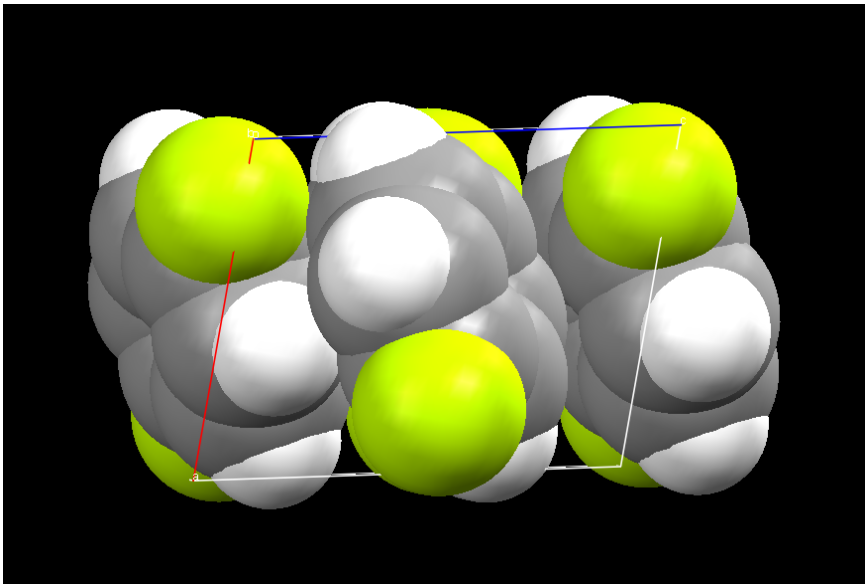


Figure 35: Crystal structure of 1,4-difluorobenzene, $P2_1/c$ Z=2.

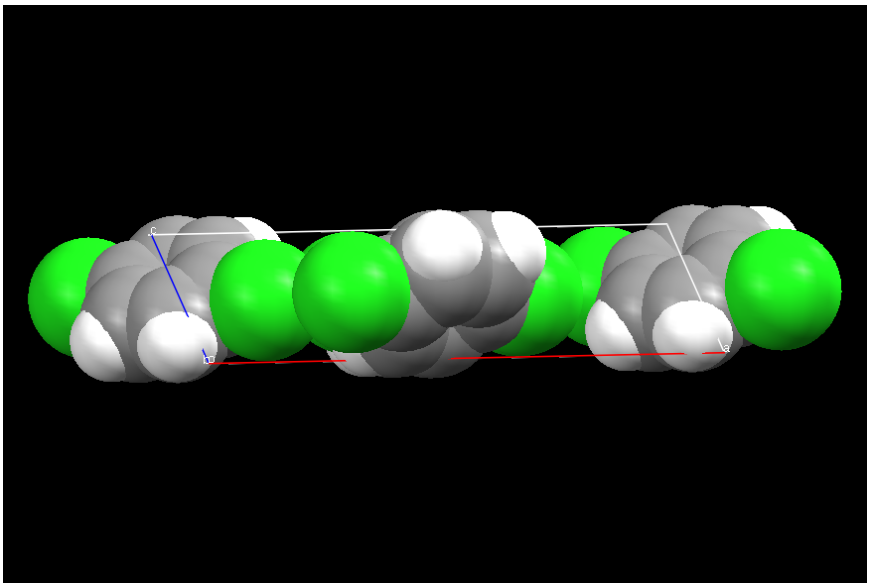


Figure 36: Crystal structure of 1,4-dichlorobenzene, P21/a Z=2.

In all these systems, there are clear examples of weak interactions in the form of hydrogen bonds, as pointed out by Glusker.[105] The C-F group, so-called "organic fluorine", does not form hydrogen bonds easily. But weak hydrogen bonds of the C-H•••X-C type can be important for the stabilization of crystal structure.[85] The C-H is known to be an hydrogen-bond donor, and C-H•••O or C-H•••Cl hydrogen bonds has been previously considered in the field of crystal engineering.[106] Therefore, even though the C-H•••F interaction is "weak", it has been shown to contribute to crystal packing in van der Waals complexes.[107]

These three molecules are comparable, both because of their symmetry and because of the weak nature of C-H•••F, C-H•••O or C-H•••Cl interactions. These species have been also previously investigated in the solid state with some of the OPiX modules.[92,108]

Using the Oprop module, we also obtain the heat of sublimation as the sum of 6-exp (Equation 37) and coulombic lattice energies. This data can be directly compared with experimental heat of sublimation as shown in Reference 98.[100] In Table 16, a comparison between theoretical and experimental heat of sublimation can be seen.

Table 17: Energetics and Properties from the OPROP module for the selected molecules (Energy in kJ/mol, volume in A**3, dipole in D).

	Volume	lattice energy	total energy	dipole	Heat of sublimation
benzoquinone	96.06	-64.9	-84.7	0	68.6
difluorobenzene	266.89	-48.2	-53.2	0	35.6
dichlorobenzene	109.56	-67.9	-69.3	0	64.9

The lattice energies obtained from OPROP are in a good agreement with experimental heat of sublimation values for benzoquinone and dichlorobenzene (less than 1 kcal/mol for both), despite the well known difficulties in describing aromatic compound.[109,110]

The difluorobenzene molecule gives a less accurate, but still quite good, result (2.8 kcal/mol of difference). Previously reported data with varying approximations as well as known error in experiment, show deviations as high as 3-4 kcal/mol.[111] It has been suggested that experimental inaccuracies are more likely to produce lower-than-exact values for the most part.[112,113] Here, we note that the experimental results are lower than the theoretical results.

We also assume when comparing calculated packing energies with observed ΔH values, that the molecule has the same conformation in the solid and in the gas phase (where the electronic structure is taken from for the procedure). In this case, the heat of sublimation should be smaller than the calculated potential energy by an amount equivalent to the gain in conformational energy in going from the crystal to the gas-phase molecule.

Using the PROM module, a variety of polymorphs were generated with the following space group: C2/c, P1-, $P2_1$, $P2_12_12_1$ and $P2_1/c$, for each of the three molecules. Of all polymorphs generated, the most stable polymorph of the group was selected and investigated by means of the Pixel module in order to evaluate the different contributions and the total energy of the polymorph. In Table 16 are reported the different contributions to the total energy for the selected polymorph of each molecules. The number of polymorphs corresponds to the final number of polymorphs predicted by PROM after the cycle of optimization. The range of energy corresponds to the minimum and maximum energy provided by PROM for the predicted polymorphs.

Of the various forces, one can see that dispersion forces play a fundamental role in the stabilization of the crystal structure like it is reported in literature.[5] It is not possible, based on the small difference in total energy between polymorphs, to define a favored space group. It is already known from experimental studies that heats of sublimation of polymorphs are known to differ by only a few percent.[114]

Table 18: Energy Contribution of crystal polymorphs using MP2/cc-pVDZ (kJ/mol).

benzoquinone	**C2/c**	**P1b**	**P21**	**P212121**	**P21/c**
coulomb	-24.1	-24.2	-24.2	-24.4	-32.5
dispersion	-74.7	-74.8	-75.6	-75	-73.7
repulsion	62.9	64.6	65.4	63.4	59.7
polarization	-14.6	-14.5	-14.7	-14.7	-10.9
total	-50.5	-48.9	-49	-50.6	-57.4
6-exp	-73.6	-73.6	-73.6	-73.7	-74.4
polymorph number	6	29	7	8	18
range	-66.7,-75.1	-64.33, -75.1	-69.4, -75.5	-62.6, -75.1	-68.9, -75.7

difluorobenzene	**C2/c**	**P1b**	**P21**	**P212121**	**P21/c**
coulomb	-	-14.5	-17.7	-13.5	-
dispersion	-	-66.4	-70.9	-65.9	-
repulsion	-	49.2	59.4	48.3	-
polarization	-	-5.9	-6.7	-5.7	-
total	-	-37.7	-35.8	-36.8	-
6-exp	-	-51.4	-51	-51.4	-
polymorph number	0	8	2	6	0
range	-	-50.8,-52.5	-50.2,-52.2	-48.2,-52.4	-

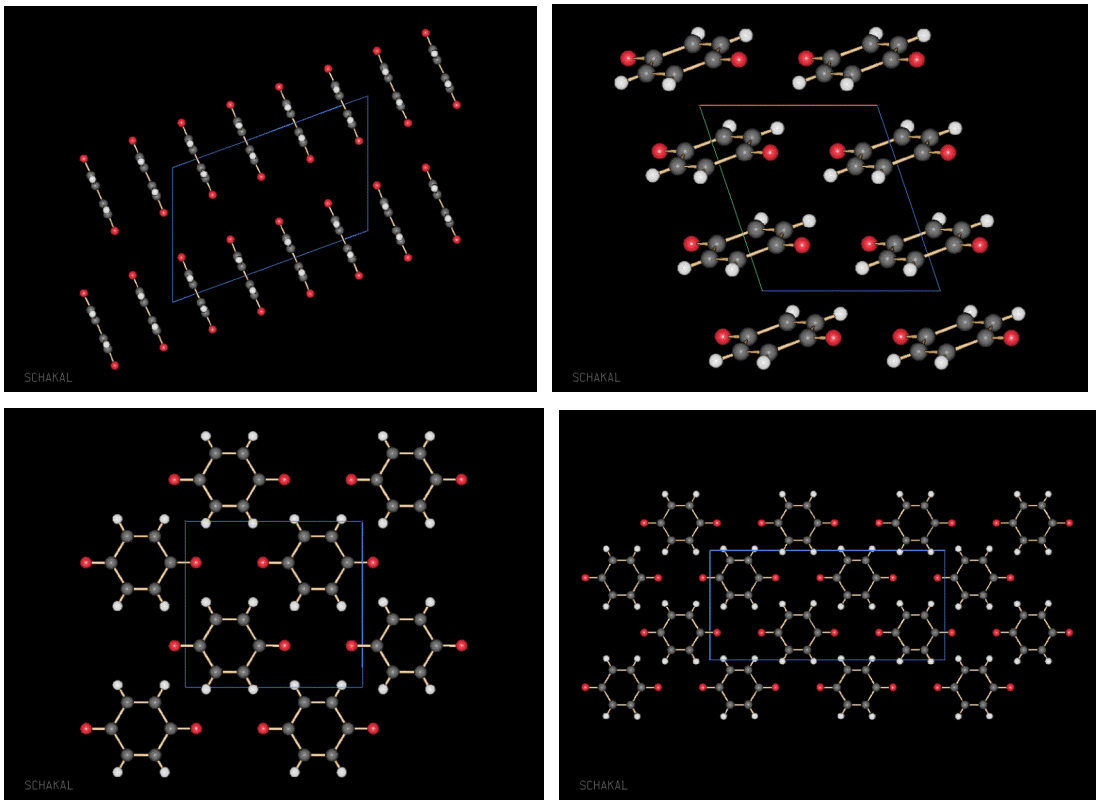
dichlorobenzene	**C2/c**	**P1b**	**P21**	**P212121**	**P21/c**
coulomb	-29.1	-24.9	-28.6	-28.1	-28.9
dispersion	-110	-115.8	-112.9	-106.7	-111.1
repulsion	81.2	98	87.4	84.9	85.2
polarization	-11.4	-13.6	-12.6	-12.4	-13.7
total	-69.4	-56.2	-66.7	-62.3	-68.5
6-exp	-71.2	-72.7	-71.9	-69.8	-71.3
polymorph number	20	18	8	6	25
range	-66.2,-74.1	-68.2,-74.1	-67.9,-73.4	-69.6,-71.5	-65.6, -72.8

For the benzoquinone, the difference in energy between the possible predicted polymorphs is less than 1 kJ/mol. The monoclinic $P2_1/c$ packing is the most stable polymorph in agreement with experimental data. As reported in previous publications, dispersion is the most important contribution for stabilization.[5,29] The second more

stable polymorph is the orthorhombic $P2_12_12_1$, followed by the monoclinic $P2_1$, the triclinic P1- and monoclinic C_2/c.

Figure 37 shows the crystal structures of the benzoquinone polymorphs predicted by the PROM module, and subsequently used with PIXEL. These representations help to visualize what interactions are favored in the crystal arrangement. The difference between the many possible polymorphs is small, and in all cases the crystal structures are held together mostly by C-H•••O hydrogen bonds.

In the $P2_1/c$, the oxygen atoms are oriented toward hydrogens of other benzoquinone, resulting in stabilizing through dispersion energy. Results of other space groups also show arrangements with oxygen facing the hydrogens. However, the repulsion forces are less important in the case of $P2_1/c$ than for the other space groups, as shown by the use of the energy decomposition in PIXEL.



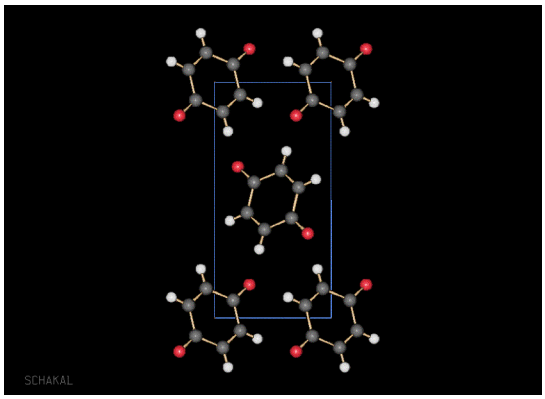


Figure 37: Crystal structures of benzoquinone predicted by PROM and used in PIXEL. (a) C2/c space group. (b) P1b space group. (c) P21 space group. (d) P212121 space group. (e) P21/c space group.

The most stable polymorph predicted for difluorobenzene is P1-. However, we do not find any polymorphs corresponding to the experimentally observed $P2_1/c$, and C2/c. There is no obvious explanation for this prediction, other than the polymorph prediction is based on empirical representations of forces, and additionally there is no constraint on the number of polymorph generated. For both of the predicted structures, dispersion effects appear to be the stabilizing force.

Figure 38 illustrates the crystal structures of the p-difluorobenzene polymorphs as predicted by the PROM module, and later used for the energy decomposition of PIXEL. These representations help us to see what interactions are favoring the crystal arrangement. The difference between the possible polymorphs is small, and in this sense similar to benzoquinone. In each case, the crystal structures appear to be held together mostly by C-H•••F hydrogen bonds.

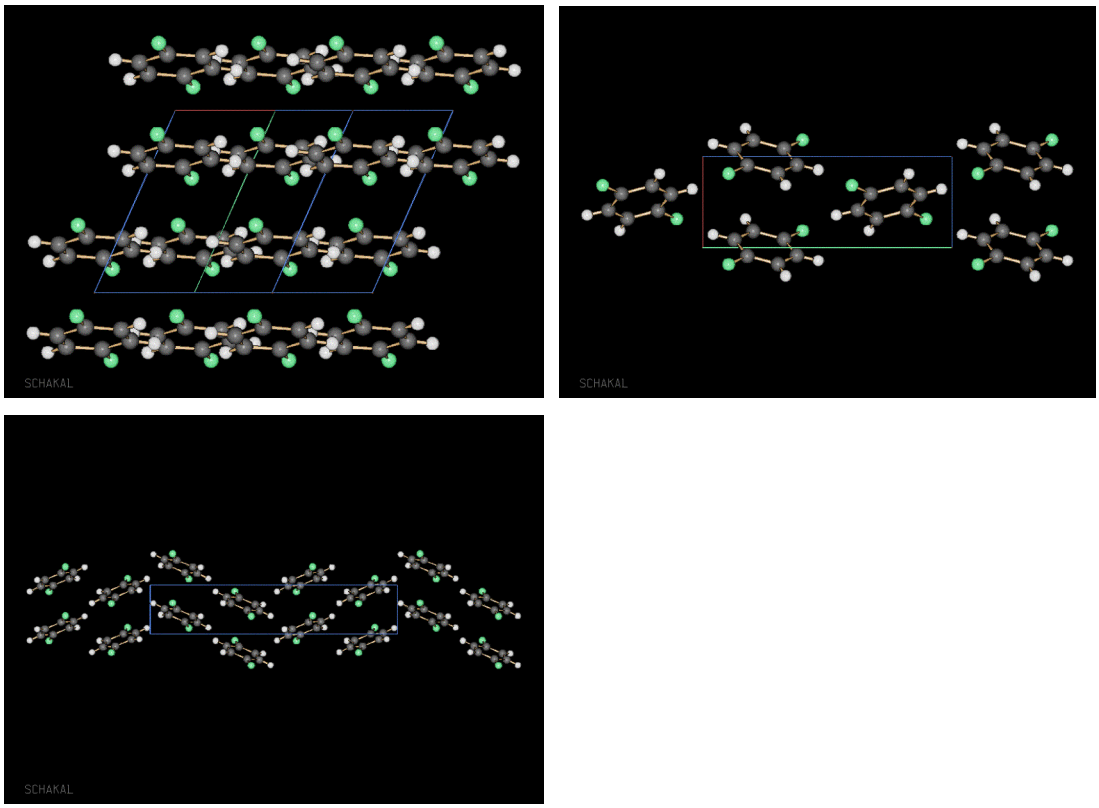
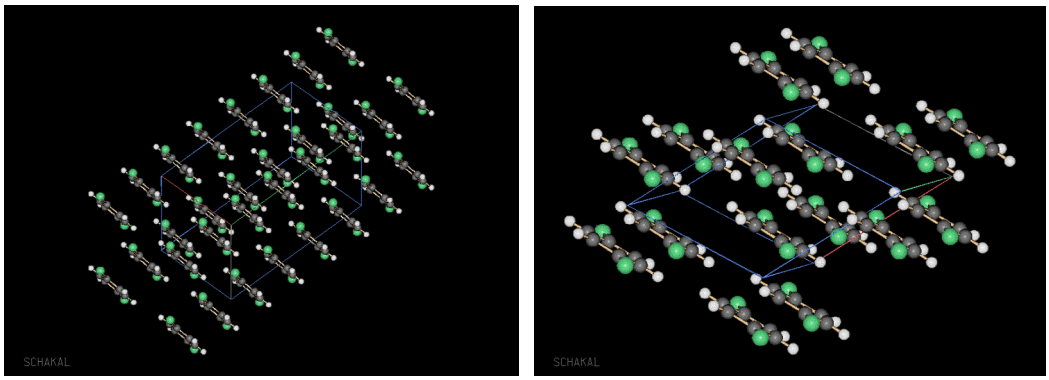


Figure 38: Crystal structures of p-difluorobenzene predicted by PROM and used in PIXEL. (a) P1b space group. (b) P21 space group. (c) P212121 space group.

The difference in energy between the polymorphs of dichlorobenzene is only about 1-2 kJ/mol. The $P2_1/a$, experimental packing is very similar to $P2_1/c$, the only difference consisting in the glide displacement direction (x instead of z for $P2_1/c$). The preferred space group according to computational analysis is also a monoclinic group.



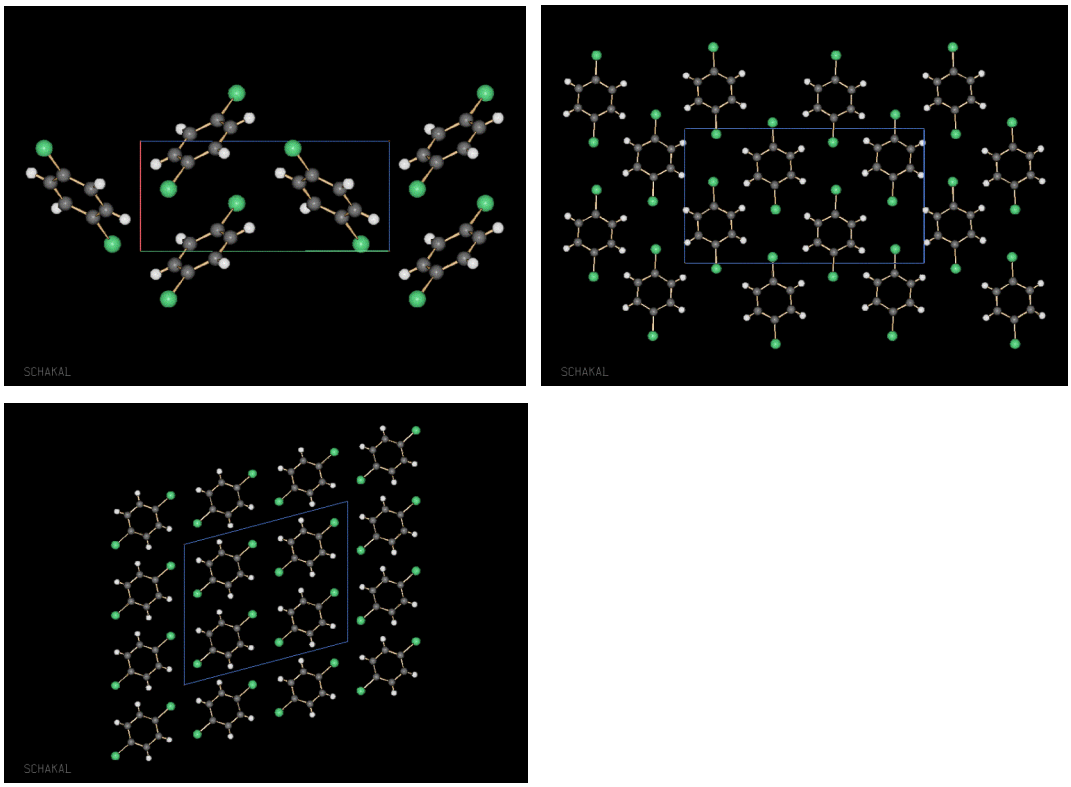


Figure 39: Crystal structures of p-dichlorobenzene predicted by PROM and used in PIXEL. (a) C2/c space group. (b) P1b space group. (c) P21 space group. (d) P212121 space group. (e) P21/c space group.

Figure 39 illustrates the crystal structures of the p-dichlorobenzene polymorphs as predicted by the PROM module, and later used with the energy decomposition module PIXEL. The difference between the possible polymorphs is also small for this third case, similar to the other two cases, again the crystal structures appear to be held together mostly by C-H•••Cl weak bonds. Stacking seems to be slightly more efficient when molecules are parallel.

The Pixel module was employed using the density cube file obtained with quantum mechanical, at different levels of theory (Table 17). We report in Table 17 the different contributions of individual energies to the total energy for the three systems above, using density cube files calculated with the M06[115] and M06-2X new hybrid meta exchange-correlation DFT functionals. The M06 functional is parameterized including both transition metals and nonmetals, whereas the M06-2X functional is a high non-local functional with double the amount of non-local exchange (2X), and is parameterized only for nonmetals. Two additional methods

used included MP2 and the hybrid DFT functional B3LYP. The basis set used is the same across the set of methods: the Dunning cc-pVDZ.[116]

Table 19: Energy contributions calculated from Pixel (kJ/mol).

MP2/cc-pVDZ	coulomb	dispersion	repulsion	polarization	total	6-exp	REFCODE	Heat of sublimation
benzoquinone	-31.7	-56.1	44	-8.6	-52.4	-63.5	BNZQUI03	68.6
difluorobenzene	-12.1	-50	25.3	-3.6	-40.4	-47.2	FACGEV	35.6
dichlorobenzene	-15.9	-79.4	40	-5.1	-60.4	-66.2	DCLBEN01	64.9

M06/cc-pVDZ	coulomb	dispersion	repulsion	polarization	total	6-exp	REFCODE	Heat of sublimation
benzoquinone	-33.9	-55.9	40.1	-8.8	-58.6	-63.5	BNZQUI03	68.6
difluorobenzene	-12.3	-49.7	23.2	-3.5	-42.2	-47.2	FACGEV	35.6
dichlorobenzene	-15.1	-78.9	36.7	-4.8	-62.1	-66.2	DCLBEN01	64.9

M06-2X/cc-pVDZ	coulomb	dispersion	repulsion	polarization	total	6-exp	REFCODE	Heat of sublimation
benzoquinone	-34	-55.9	41.7	-8.8	-57	-63.5	BNZQUI03	68.6
difluorobenzene	-13	-49.6	24.3	-3.7	-42	-47.2	FACGEV	35.6
dichlorobenzene	-15.8	-79	38.1	-5	-61.6	-66.2	DCLBEN01	64.9

B3LYPcc-pVDZ	coulomb	dispersion	repulsion	polarization	total	6-exp	REFCODE	Heat of sublimation
benzoquinone	-33.6	-55.9	43.4	-8.9	-55	-63.5	BNZQUI03	68.6
difluorobenzene	-12.1	-49.7	25.1	-3.5	-40.2	-47.2	FACGEV	35.6
dichlorobenzene	-15.5	-79	38.1	-5	-61.3	-66.2	DCLBEN01	64.9

The results across the set of methods are very similar, and in each case the M06-2X functional appears to be giving the largest total energies compared to the other methods. In all three cases, dispersive forces appear to be the largest component responsible for the cohesion of the crystal. It is particularly important for the dichlorobenzene because of the chlorine-chlorine interaction (induced dipole). The polarization effect for benzoquinone is important due the oxygen atoms acting on the surrounding molecules. In the case of chlorobenzene, the Cl•••Cl interactions play a significant role in the stabilization, mainly because of dispersion.[84,92]

4.4 Conclusion

Starting from experimental data, a methodology to study crystal environment was presented in this chapter. A recent method for the evaluation of intermolecular potentials, based on semi-classical density sums (SCDS), has been described and applied to prototypical aromatic systems by means of the OPiX package. Application of this method to organic crystals to define the different contributions of energetic forces, as well as to evaluate the lattice energy, is illustrated. Applications of the methodology to three well-known benzoic systems as presented.

Using this package, a more clear understanding of how molecules pack in a crystalline structure, and what the primary forces involved are, can be obtained. In all presented cases, dispersion is the main stabilizing force as previously reported.[4,5,117] A method for generating polymorph structures is demonstrated, and shown to provide good agreement with experimentally known data.

5 Applications of Methodology: Functionalized Corannulene Systems

5.1 Introduction

Polycyclic aromatic hydrocarbons (PAHs) constitute a diverse class of organic chemical compound. These widely varying structures all consist of some number of fused aromatic rings, which sets up a network of delocalized electrons, giving the structure characteristic reactivity. Many such examples of PAHs are found in the environment, including oil, coal, tar, which are sources of PAH. PAHs are also byproducts of fuel burning (fossil fuel or biomass), and can be found in comets and meteorites. PAH have even been cited in a hypotheses related to their involvement in a pre-RNA world basis for the origin of life.[118] PAHs are also used in the plastic industry and for synthetic fiber, but can be found in air, water and some foods also.

The discovery of carcinogenic and mutagenic properties of a few polyarenes[119] enhanced the research in the area of PAH in the mid-1930s.[120] However, the major increase of interest in PAH's happened during the 1980s when fullerenes were discovered by Kroto, Curl and Smalley in 1985, research that earn them the Nobel Prize in Chemistry in 1996.[121] This new family of carbon allotropes brought a new life to chemistry of PAH.

According to International Union on Pure and Applied Chemistry (IUPAC), the simplest PAHs are phenanthrene and anthracene, since benzene and naphthalene are not part of the class even though they are chemically related. The most common PAH contains five or six member rings, but the most famous PAH is buckminsterfullerene (or (C60-Ih)[5,6]fullerene), discovered by laser vaporization of graphite in 1985[121] and named after the architect R. Buckminster Fuller[121] who pioneered the use of geodesic domes. This discovery opened a new domain of research into cage-like molecules in the shape of spheres, ellipsoids, and tubes, and since then, a large number of investigations have been focused on fullerenes. This unique motif is the skeleton for a class of curved PAHs, whose molecular structure corresponds to fraction of the full fullerene. In particular, two molecular types within this class have been synthesized, $C_{36}H_{12}$, and corannulene ($C_{20}H_{10}$).[122] This second molecule, corannulene, is the focus of the present work.

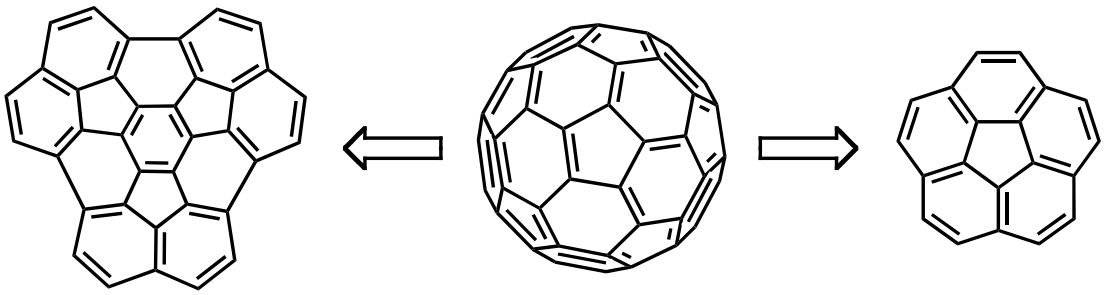


Figure 40: Buckminsterfullerene (2, middle) and Corannulene (1, right).

5.2 Corannulene Background

Corannulene, or dibenzo[ghi,mno]fluoranthene, consists of a cyclopentane ring fused together with five benzene rings. Corannulene is considered as a fragment of buckminsterfullerene, as the curved geodesic polyarene face of the full ball structure. Due to its curvature, the molecule is also referred to as a buckybowl, analogous to the reference of buckyball for buckminsterfullerene. Corannulene was first synthesized in 1966 by Lawton and Barth using a 17-step procedure,[122] with a very low overall yield (1%), but with a successful crystal structure[123] that confirmed curvature of the molecule. The crystal structure enabled prediction of the bond length and angles of the molecule, which were later confirmed by electron diffraction.[124] The bowl dynamics was, however, not determined until later studies[125] when new more robust synthesis became available.[126-128]

5.3 Structure and Properties

X-ray crystallography of corannulene reveals a bowl depth of 0.875 Å, as measured from the plane formed by the central cyclopentane base to the edge carbons of the bowl. In buckminsterfullerene, this same distance is 1.50 Å, showing a much deeper curvature. Corannulene has four different carbon-carbon bond lengths: rim, flank, spoke and hub, as depicted in Figure 41, as compared to C_{60}, which has only 2 different bond lengths.

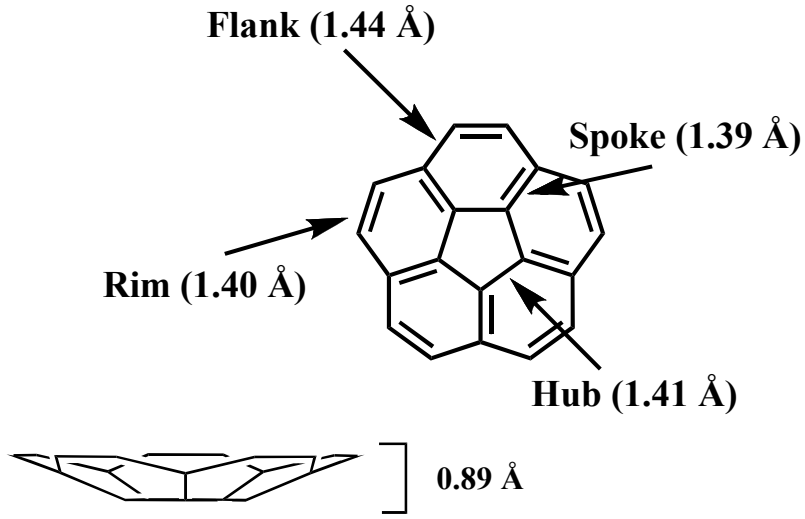


Figure 41: Structure of Corannulene**.**

Another important structural measure of the non-planarity of corannulene is the π-Orbital Axis Vector (POAV), first proposed by Haddon.[129] The POAV angle is defined as shown in Figure 42, by taking a perpendicular vector at any one of the central carbon atoms, and measuring the angle from that vector to one of the associated carbon bonds. If the angle between the POAV and the σ-bonds is $\theta_{\sigma\pi}$, then the POAV angle would be $\phi_{POAV} = \theta_{\sigma\pi} - 90$. Typically, a sp2 hybridized carbon atom forms a planar σ-framework with its neighbors, so the p-orbital is perpendicular to the σ-bonds. For example, the carbons of graphite have $\phi_{POAV} = 0^{\circ}$, and the carbon of buckminsterfullerene all have $\phi_{POAV} = 11.6^{\circ}$ [130]. The POAV angles of corannulene are respectively 8.7° (hub), 5.5° (spoke) and 1.1° (rim). The POAV is a measure of the curvature-induced local weakening of π-conjugation.[129,131,132] As such, the reactivity of buckminsterfullerene is expected to be greater than corannulene.[128]

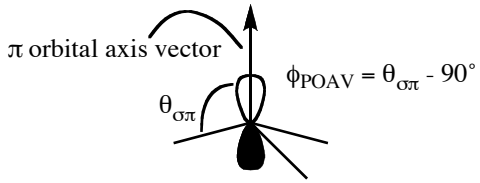


Figure 42: Definition of POAV angles.

PAH composed of only six member rings are said to be alternant hydrocarbons, for example as found in coronene (Figure 43).[120]

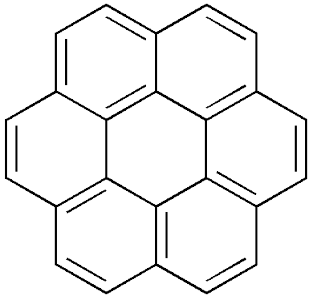


Figure 43: Coronene structure ([6]circulene).

Corannulene is a non-alternant PAH, presented by Barth and Lawton as an annulene-within-an annulene model of aromaticity. In such a model, an aromatic cyclopentadienyl anion (6 electrons / 5 carbon atoms) is surrounded by an aromatic (14 electrons / 15 carbon atoms) annulenyl cation. Even though the nonplanar 20 π-electron structure fails the Hückel rule of 4n+2, a polar version of the molecule (Figure 44) would contribute to the π electron distribution of the system.[133]

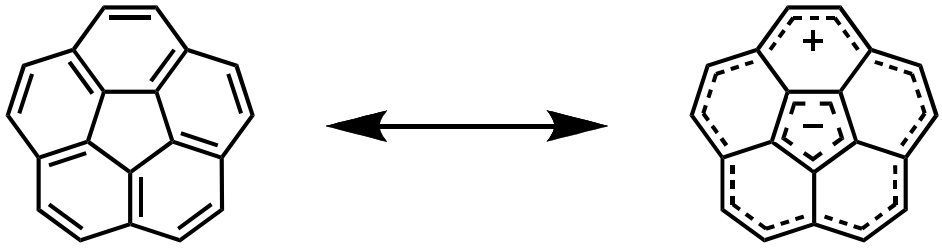


Figure 44: Resonance structure of Corannulene.

5.3.1 Bowl to Bowl Inversion

Lawton and Barth first suggested that the bowl and planar conformation were respectively minima and maxima on the corannulene potential energy surface (PES), giving rise to its dynamic behavior. This bowl-to-bowl inversion process is depicted in Figure 45. However, at that time, they were unable to prove or measure this dynamic behavior experimentally.

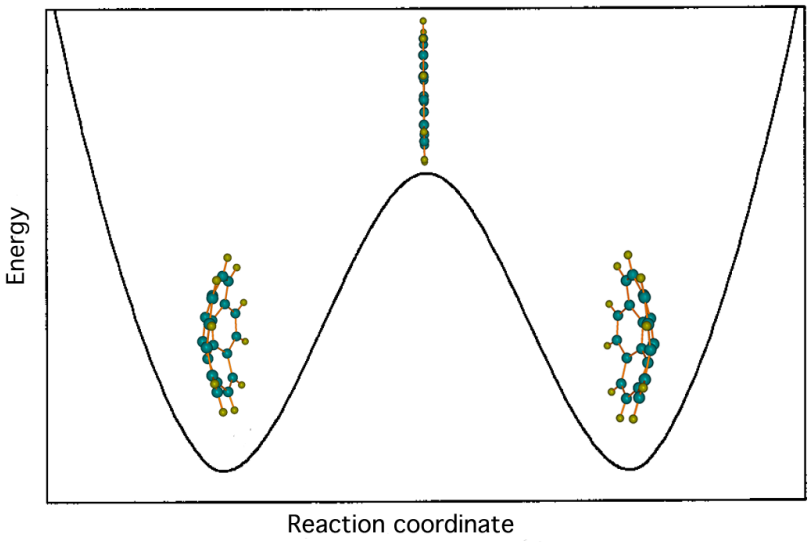


Figure 45: Bowl to bowl process of Corannulene.

Unfortunately, the highly symmetric C_{5v} structure prevents direct measurement of the interconversion barrier by NMR techniques. In order to overcome this problem, several derivatives incorporating stereochemical probes have been synthesized, enabling the corannulene interconversion barrier to be measured.[125] In that same study, computational methods were successfully used for the first time to quantify the interconversion barrier. This was beneficial, since the experimental procedure only enabled an estimate of the barrier, since the use of probes actually perturbs the barrier that is measured. Experimentally, the barrier was estimated at 11.5 kcal/mol (± 0.6 kcal/mol) by plotting the effect of sequential bromomethylation on the barrier. This was then compared to the theoretical determination, the latter of which is further elaborated on in a next chapter. Subsequently, both theory and experimental techniques were used to study a large series of corannulene derivatives, enabling a predictive structure-energy correlation between bowl depth and inversion barrier.[125]

5.3.2 Theoretical Treatments of Corannulene

Theoretical investigations of corannulene in this research groups has revealed a sensitivity of method in determining, in particular, curvature and dynamical properties of benzene. While structural parameters, such as bond distances and angels, show in general a good correlation with experiment, dynamic properties are shown to be much more difficult to predict within experimental accuracy. In

particular, issues of wave function type and basis set, become important determinants to how good the predictions of corannulene structure and associated properties will be. As such, part of the current research involved a full basis set investigation of corannulene as well as functionalized corannulenes.

Structural computations of all corannulene compounds were performed using the following wavefunction types: restricted Hartree-Fock (RHF), Møller-Plesset perturbation theory (MP2) and hybrid Density Functional Theory (HDFT), using the GAMESS package. The HDFT method employed Becke's 3-parameter functional[50] in combination with non-local correlation provided by the Lee-Yang-Parr expression[53] that contains both local and non-local terms, B3LYP. Although it has already been shown that these levels of theory are appropriate to obtain an accurate description of the structure, as coupled with the Dunning's correlation consistent basis set (cc-pVDZ), a [3s2p1d] contraction of a (9s4p1d) primitive set, a full suite of basis sets, in combination with the different wave function types, was investigated to better understand the sensitivities associated with the curvature, higher order properties, and dynamical properties (e.g., barrier to interconversion).

All structural parameters, energetics, as well as some associated properties were all calculated for the lowest energy geometry (bowl) as well as the transition state (flat) of corannulene. These results were compared to available X-ray crystal structures when available. As it is mentioned in chapter 2, molecular structure determined by X-ray analysis are affected by crystal packing forces, which can distort molecules significantly from their true geometries they may adopt in the absence of those influences. On the contrary, *ab initio* calculations deal, in general, with isolated molecules.

For each structure, a vibrational analysis was performed to characterize the calculated structures as true minimum in the case of the ground state, or transition state structure, as in the case of the flat structure. Other calculated properties included the highest occupied molecular orbital, (HOMO), which provides a crude estimate of the Ionization potential via Koopmans Theorem approximation[82], the lowest unoccupied molecular orbital, (LUMO), and therefore, the HOMO/LUMO gap, the dipole, and the inversion barrier of corannulene with and without zero point correction (Table 20).

Table 20 and Figure 46 presents a summary of geometric data as well as a variety of properties, for both ground and transition state, as a function of

wavefunction type and basis set. Calculations were performed using the following wavefunction types: restricted Hartree-Fock (RHF), Møller-Plesset perturbation theory (MP2) and hybrid Density Functional Theory (HDFT), and a variety of basis set.

Structural information using RHF or DFT methods, in particular, bond length and bond angles, is able to closely represent crystallographic data. However, information related to the curvature of the molecule, such as bowl depth and POAV angle, demands at least DFT level of theory for accurate results.

Petrukhina[134] reported the use of B3LYP/6-31G* for structural analysis, which is much more limited in basis set flexibility than is the cc-pVDZ, however as expected for simple structure (bond lengths and angles) provides similar level of prediction to that of to B3LYP/cc-pVDZ. Their particular study focuses mainly on basic structural information such as bond length, angles, but not the energetics and structure related to the curvature of the molecule, for which this method would have more difficulty. Most standard density functionals, including B3LYP, are not able to accurately predict the bowl depth and curvature of the molecule, and consequently, the inversion barrier (e.g., see Table 19). This is primarily due to the limitations in these functionals ability to capture dispersion effects. Some of the more modern density functionals, such as DFT-D,[135] include an evaluation of dispersion effect, and are therefore much more accurate in their prediction of the more difficult parameters (e.g., dipole, bowl-depth, curvature) and also the bowl-to-bowl inversion energy.[21]

In terms of predicted barriers, both B3LYP/6-31G(d) nor B3LYP/cc-pVDZ underestimate the true barrier, as might be expected from the error in predicted bowl depth in particular. Figure 46 summarizes the variation in the inversion barrier as a function of wavefunction type and basis set. It shows the variability in the inversion barrier depending on the wavefunction type and basis set.

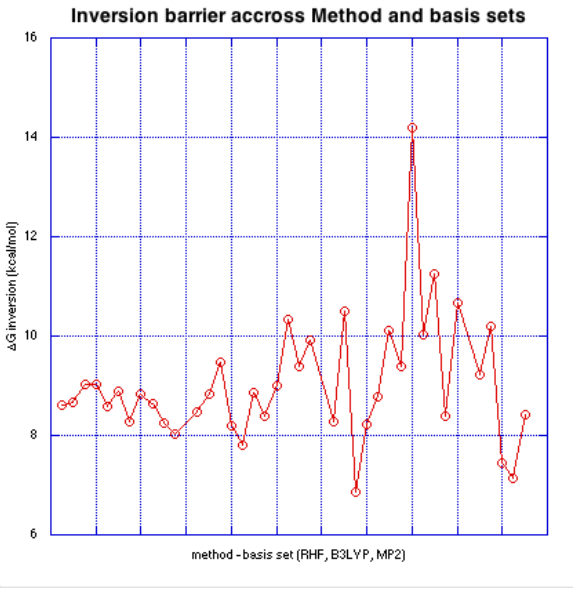


Figure 46: Inversion barrier (kcal/mol) (y axis) (ZPE corrected) of corannulene as function of wavefunction type (x axis).

Despite this relatively poor performance in prediction of interconversion barrier, before the development of dispersion enabled DFT, our group showed that the use of MP2 theory on top of the B3LYP geometries can lead to a good estimate of the barrier height. In particular, as shown,[136] the MP2/cc-pVDZ//B3LYP/cc-pVDZ, is one of the most accurate level of theory for the consistent prediction of inversion barrier of corannulene, as well as functionalized corannulenes. More recently, our group has implemented a series of dispersion corrected DFT functionals, which are not only able to describe accurately the structural properties of corannulene but are also able to predict accurately this barrier to interconversion.

Table 20: Basis set study of corannulene. ΔG in kcal/mol. dipole in Debye, HOMO-LUMO in eV, bowl depth and length in Å, and frequency in cm-1. (MP2/-//B3LYP/DZV(d,p) means MP2/DZV(d,p) single point using an optimized geometry at B3LYP/DZV(d,p)).

	ΔG	ΔG with ZPE	dipole bowl	HOMO bowl	LUMO bowl	HOMO flat	LUMO flat	imaginary mode	bowl depth low	bowl depth high	rim bond bowl	flank bond bowl	Hub bond bowl	Spoke bond bowl	rim bond flat	flank bond flat	Hub bond flat	Spoke bond flat
RHF/STO-3G(2d,p)	8.77	8.61	1.02	-6.36	4.83	-6.51	4.78	118.92	0.48	0.81	1.36	1.46	1.42	1.36	1.37	1.47	1.41	1.35
RHF/6-31G	8.86	8.68	2.25	-8.08	1.85	-8.11	1.79	116.44	0.49	0.82	1.38	1.45	1.41	1.36	1.39	1.46	1.39	1.35
RHF/6-31+G	9.48	9.04	2.44	-8.25	1.44	-8.24	1.38	119.00	0.49	0.83	1.38	1.45	1.41	1.36	1.39	1.46	1.39	1.35
RHF/6-31++G	9.38	9.02	2.45	-8.25	1.07	-8.24	1.09	118.08	0.49	0.83	1.38	1.45	1.41	1.36	1.39	1.46	1.39	1.35
RHF/6-31G(d)	8.85	8.59	1.90	-7.98	1.90	-7.95	1.78	116.94	0.49	0.83	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.35
RHF/6-31+G(d)	9.53	8.89	2.24	-8.20	1.49	-8.17	1.42	119.11	0.49	0.83	1.37	1.45	1.41	1.36	1.39	1.46	1.39	1.35
RHF/6-31++G(d)	8.67	8.28	2.29	-7.76	1.09	-7.50	0.98	118.76	0.50	0.83	1.37	1.45	1.41	1.36	1.39	1.46	1.40	1.35
RHF/6-31+G(d,p)	9.50	8.85	2.28	-8.19	1.48	-8.16	1.40	118.81	0.49	0.83	1.37	1.45	1.41	1.36	1.39	1.46	1.40	1.35
RHF/6-31G(d,p)	8.90	8.64	1.89	-7.99	1.93	-7.97	1.81	116.64	0.49	0.83	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.35
RHF/6-31G(2d,p)	8.48	8.25	1.76	-7.93	1.99	-7.88	1.88	114.70	0.49	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/6-31G(3df,3p)	8.28	8.03	1.90	-8.01	1.84	-7.96	1.75	111.53	0.48	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/6-31G(3df,3pd)	8.86		1.96	-8.08	1.75	-8.03	1.67		0.49	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/6-311G	8.49	8.48	2.34	-8.22	1.63	-8.21	1.52	112.95	0.49	0.82	1.37	1.45	1.41	1.36	1.39	1.46	1.39	1.35
RHF/6-311G(d,p)	8.96	8.82	2.18	-8.14	1.65	-8.10	1.55	115.77	0.49	0.83	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/6-311G(2d,p)	9.69	9.46	2.12	-8.08	1.70	-8.01	1.62	121.02	0.50	0.84	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/6-311G(2df,2dp)	8.77	8.20	2.06	-8.13	1.74	-8.08	1.65	117.62	0.49	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/DZV(2d,p)	7.82	7.82	2.11	-8.04	1.42	-7.96	1.33	107.95	0.49	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
RHF/cc-pVDZ	9.03	8.86	2.08	-8.12	1.72	-8.09	1.62	116.05	0.50	0.83	1.37	1.45	1.41	1.36	1.39	1.47	1.40	1.35
RHF/cc-pVTZ	8.91	8.40	2.11	-8.09	1.39	-8.01	1.29	117.88	0.49	0.82	1.37	1.45	1.41	1.36	1.38	1.46	1.39	1.34
B3LYP/cc-pVDZ	9.04	9.00	1.88	-6.11	-1.57	-6.05	-1.67	108.91	0.52	0.87	1.39	1.45	1.42	1.39	1.41	1.46	1.40	1.37
MP2/-/B3LYP/cc-pVDZ	10.38	10.33	2.29	-8.04	1.52	-7.97	1.40											
B3LYP/cc-pVTZ	9.93	9.39	2.05	-6.18	-1.87	-6.10	-1.97	115.49	0.52	0.88	1.38	1.44	1.41	1.38	1.40	1.46	1.39	1.36
MP2/-//B3LYP/cc-pVTZ	10.46	9.92	2.35	-8.10	1.44	-8.02	1.35											
B3LYP/cc-pVQZ	10.26		2.04	-6.21	-1.45	-6.13	-1.52	149.35	0.52	0.88	1.38	1.44	1.41	1.38	1.40	1.46	1.39	1.36
B3LYP/DZV(d,p)	8.63	8.27	2.28	-6.06	-1.38	-5.99	-1.45	103.73	0.53	0.88	1.40	1.45	1.42	1.39	1.41	1.47	1.40	1.37
MP2/-//B3LYP/DZV(d,p)	10.84	10.49	2.52	-8.10	1.05	-8.05	0.98											

B3LYP/DZV(2d,p)	6.86	6.86	2.03	-6.01	-1.41	-5.93	-1.50	95.81	0.51	0.84	1.40	1.45	1.41	1.38	1.41	1.46	1.40	1.37
MP2/-//B3LYP/DZV(2d,p)	8.23	8.23	2.23	-8.04	1.11	-7.96	1.04											
B3LYP/DZV(2df,2p)	9.14	8.77	2.09	-6.09	-1.40	-6.02	-1.47	110.55	0.52	0.87	1.39	1.44	1.41	1.38	1.40	1.46	1.39	1.36
MP2/-//B3LYP/DZV(2df,2p)	10.49	10.11	2.32	-8.09	1.13	-8.01	1.05											
B3LYP/TZV(d,p)	10.64	9.39	2.26	-6.21	-1.43	-6.13	-1.51	119.55	0.53	0.89	1.39	1.45	1.42	1.38	1.40	1.46	1.40	1.36
MP2/-//B3LYP/TZV(d,p)	15.43	14.19	2.54	-8.12	1.07	-8.07	1.00											
B3LYP/TZV(2d,p)	10.72	10.04	2.16	-6.17	-1.48	-6.10	-1.54	117.81	0.53	0.89	1.38	1.44	1.41	1.38	1.40	1.46	1.39	1.36
MP2/-//B3LYP/TZV(2d,p)	11.94	11.25	2.47	-8.10	1.09	-8.03	1.02											
B3LYP/DZP	8.72	8.38	2.35	-6.04	-1.36	-5.96	-1.43	102.40	0.53	0.88	1.40	1.45	1.42	1.39	1.41	1.47	1.40	1.37
MP2/-//B3LYP/DZP	11.02	10.68	2.58	-8.08	1.06	-8.03	0.99											
B3LYP/6-31G(3df,3pd)	10.08		1.90	-6.11	-1.22	-6.04	-1.29		0.53	0.88	1.39	1.45	1.42	1.38	1.40	1.46	1.40	1.36
B3LYP/6-311G(2df,2pd)	9.73	9.22	1.99	-6.18	-1.26	-6.11	-1.33	114.94	0.52	0.87	1.38	1.44	1.41	1.38	1.40	1.46	1.39	1.36
MP2/-//B3LYP/6-311G(2df,2pd)	10.72	10.21	2.30	-8.05	1.21	-7.97	1.14											
MP2/cc-pVDZ	10.37	7.44	2.22	-6.27	1.18	-6.22	1.09	115.48	0.55	0.90	1.40	1.45	1.42	1.40	1.42	1.46	1.40	1.38
MP2/DZV(2d,p)	8.29	7.14	2.29	-6.24	1.04	-6.16	0.95	97.34	0.54	0.88	1.40	1.45	1.42	1.39	1.42	1.46	1.40	1.38
MP2/DZV(d,p)	10.84	8.42	2.51	-6.29	0.98	-6.23	0.90	116.45	0.55	0.90	1.40	1.45	1.42	1.40	1.42	1.46	1.40	1.38

5.3.3 Substituted Corannulene

5.3.3.1 Background

There are two different classes of substitutions on corannulene that have been studied (Figure 47). One class involves functionality that results in repulsive interactions at the rim, causing a flattening out of the bowl and a decrease in the interconversion barrier (Class I). The second class involves annelated functionality, the result of which is to cause further restriction in the ability of the bowl to interconvert, thereby increasing ΔG_{inv} (Class II). Class I type interactions involve substituents with sterically demanding groups, while class II substituents involve annelations at the edge of the bowl structure.[125]

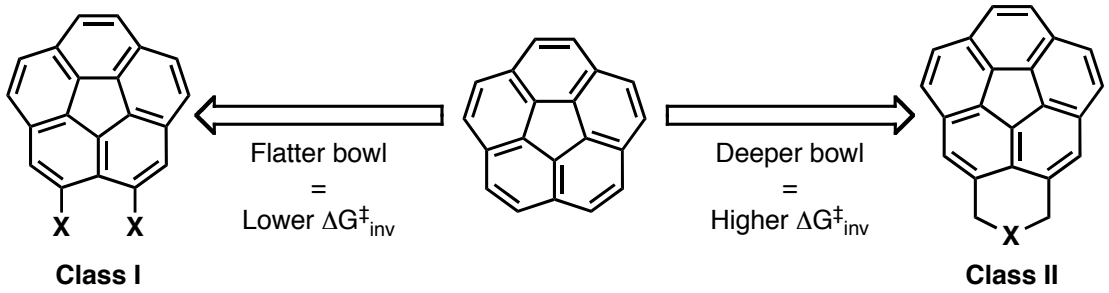


Figure 47: Class of substituents of corannulene.

Previous investigations across a large series of functionalized corannulene derivatives from each of these classes has allowed determination of a correlation between bowl depth and interconversion energy.[137,138] This correlation has been shown to be an efficient way to describe stereo-electronic effects for corannulene. Figure 48 shows the resulting functional fit for a series of corannulene derivatives derived from each of Class I and Class II, for both experimental as well as theoretical (MP2/cc-pVDZ//B3LYP/cc-pVDZ) results.

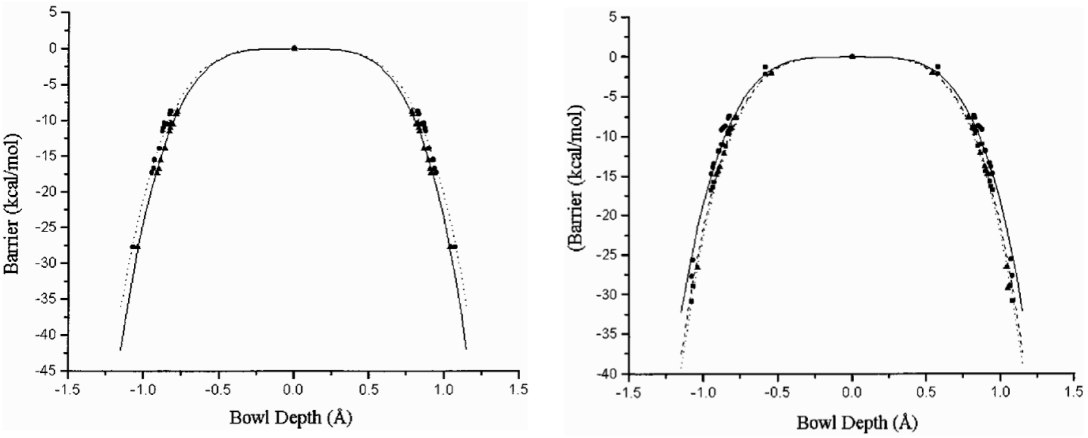


Figure 48: Experimental (right) and theoretical (left) structure-energy correlation of inversion barrier versus bowl depth.

These results show that the bowl-to-bowl inversion process is driven by changes in stability of the ground state with respect to the transition state, depending on the substituent effect. In the cases where the ground state is destabilized by repulsive interactions, one will observe a decrease in the activation energy of the bowl-to-bowl dynamical process. In contrast, substituents that produce a stabilizing effect in the ground state relative to the transition state, results in an increase in inversion barrier. The reaction path profile for corannulene is illustrated in Figure 49 by two symmetric minima connected through the local maxima of the planar transition state. The correlation relationship, made by Seiders et al,[125] revealed a quartic dependence between structure - equilibrium bowl depths (x_{eq}) - and energy inversion barriers (ΔG), as shown in Equation 38.

$$E = ax^4 - bx^2 \qquad \textbf{38a}$$

$$dE/dx = 4ax^3 - 2bx = 0 \qquad \textbf{38b}$$

$$x = 0 \quad b = 2a(x)^2 \qquad \textbf{38c}$$

$$\Delta E = E(x_{eq}) - E(x_0) = a(x_{eq})^4 - 2a(x_{eq})^2(x_{eq})^2 = -a(x_{eq})^4 \qquad \textbf{38d}$$

Equations 38a-d describes the set of reaction profiles, including the parent corannulene, through the extrema of maximum and minimum barrier, which leads to

the quartic relationship between inversion barriers with the bowl depth. The coefficient *a* is a constant among derivatives, and *b* is a variable that accounts for the particular shape of the double-well potential. When *b* is large, the double well potential is deeper and the inversion barrier larger than in the parent, corannulene. When *b* is small, the double well potential is more shallow and the barrier smaller, than in the parent corannulene. Correlation between inversion barriers and bowl depth, obtained either experimentally or by calculation, represents in general the possible perturbations of corannulene through the effects of substituents.

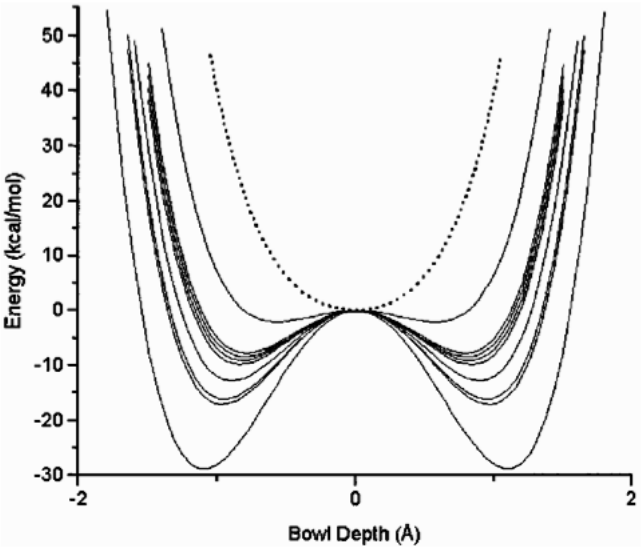


Figure 49: Reaction profiles representing the perturbation of bowl-shaped corannulene derivatives (solid lines, double-well potentials) toward flat corannulene derivatives (dashed lines, single-well potential).

5.3.3.2 Penta-substituted corannulene derivatives

Once such a correlation as presented in 5.3.3.1 is found for a particular molecular system, it becomes very interesting to observe any deviations from that general correlation. Such a deviation was observed with a special penta-substitution pattern. Several penta-substituted corannulene derivatives, studied both experimentally as well as theoretically (MP2/cc-pVDZ//B3LYP/cc-pVDZ level of theory), revealed a particular derivatives **3** and **4** of particular interest due to observed higher inversion barrier than predicted by the general correlation.

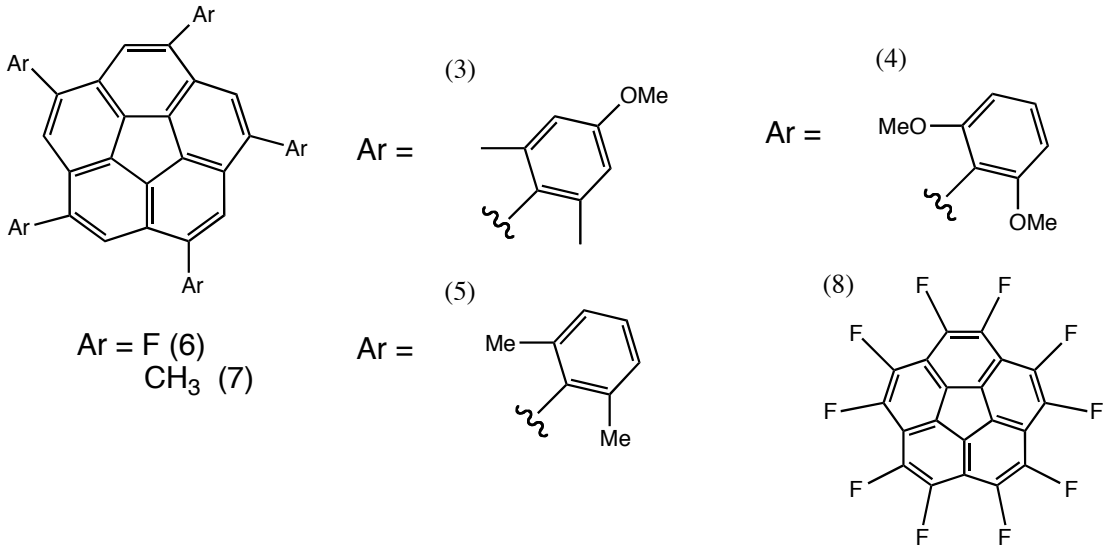


Figure 50: Substituted penta-substituted corannulene derivatives: penta(dimethyl methoxyphenyl)corannulene (3) and penta(dimethoxyphenyl)corannulene (4), penta(dimethylphenyl)corannulene (5), pentafluorocorannulene (6), pentamethylcorannulene (7). decafluorocorannulene (8).

The results from previous computational analysis, as well as the present, for penta-substituted corannulenes **3-8**, are presented in Table 20. The previous study[139] used M06-2X/cc-pVDZ method (a more modern DFT functional containing dispersion correction[115]), while the present analysis includes B3LYP/cc-pVDZ and MP2/cc-pVDZ. The symmetry of the ground state in these systems is C_5. The structural results are also compared to crystal structure results from experiment, which is slightly distorted away from C_5, due to crystal packing forces.

The results presented in this work are consistent with experimental results, and inversion barrier presented here are also consistent with previously reported results. Geometries from the penta-substituted corannulene are consistent with the [5]radialene pattern observed in corannulene. The calculated values for bond length and bon angles are consistent with previously published results using electron correlation.

The distance $Me_{(endo)}/Me_{(endo)}$ for **3** is 4.64 Å, and the reported value in crystal structure range from 3.9 to 5.2 Å depending on soft potentials, the calculated depth is 0.86 Å (depth $_{hub\text{-}to\text{-}rim}$), and the observed one is 0.91 Å, meaning that the B3LYP calculation miss some effect. M06 calculation showed a bowl depth of 0.94 Å. The X-ray and the dispersion consideration suggest a deeper bowl than a simple

hybrid method such as B3LYP. The distance $Me_{(exo)}/Me_{(exo)}$ is 6.33 Å, and the distance $Me_{(endo)}/Me_{(endo)}$ across the ring is 7.51 Å. The distance $Me_{(endo)}/Me_{(endo)}$ of the flat transition state is 5.67 Å and the distance across the ring is 9.18 Å.

The distance $Me_{(endo)}/Me_{(endo)}$ for **4** is 5.74 Å, the distance $Me_{(exo)}/Me_{(exo)}$ is 7.17 Å, and the distance $Me_{(endo)}/Me_{(endo)}$ across the ring is 9.29Å. The bowl depth is 0.84 Å. The transition flat form of **4** present a distance $Me_{(endo)}/Me_{(endo)}$ of 6.38 Å, and a across distance $Me_{(endo)}/Me_{(endo)}$ of 10.33 Å.

The molecule **5** has also been studied using the functional M06. In this work, the calculated distance $Me_{(endo)}/Me_{(endo)}$ is 4.62 Å while the $Me_{(exo)}/Me_{(exo)}$ is 6.32 Å and the distance across ring is 7.42Å. The bowl depth is calculated at 0.85Å. The transition flat form of **5** present a distance $Me_{(endo)}/Me_{(endo)}$ of 5.67 Å, and a across distance $Me_{(endo)}/Me_{(endo)}$ of 9.17 Å.

The pentafluorocorannulene, **6**, the distance F/F is 5.21 Å while the H/H distance is 5.01 Å, the bowl depth is 0.88 Å and the cross distance is 8.49 Å. The transition state of the molecule presents a distance F/F of 5.47 Å, a H/H of 5.22 Å and a cross distance of 8.86 Å.

Compared to the decafluorocorannulene, **8**, where the distance between two fluorine atom in position 1 to 3 is 5.29 Å, the cross distance is 8.56 Å and the bowl depth is 0.83 Å. The flat transition state presents a F_1/F_3 distance is 5.48 Å and the cross distance is 8.87 Å.

The pentamethylcorannulene, **7**, presents a Me/Me distance of 5.5 Å, a cross distance of 8.90 Å while the bowl depth is 0.85 Å. The flat pentamethylcorannulene has a distance Me/Me of 5.70 Å and a cross distance of 9.23 Å. When using MP2/cc-pVDZ, the Me/Me distance is 5.48 Å and the cross distance is 8.87 Å, this tells us that for the molecule **7** the use of electron-electron correlation is not necessary for accuracy. It is the same for the flat molecule, the MP2/cc-pVDZ, the Me/Me distance is 5.71 Å and the cross distance is 9.24 Å.

In the case of the molecules **3** and **5**, the proximity of the endo methyls group in the bowl form are within the van der Waals distance between two methyl group (ca. 4.5 Å), which would explained the unusual barrier and the need to use a functional that considers dispersion effects, so the extra stabilization of the bowl form would be accurately predicted. As the results of the previously cited references pointed out, the Me/Me distance would be shorter than the sum of the van der Waals

radii, and then the ground state of the molecule would be holding the methyls of the substituent inside this range to overcome the repulsion between the methyls groups.

Table 21: Energetics and Properties of selected substituted corannulene: ΔG in kcal/mol. dipole in Debye, HOMO-LUMO in eV, bowl depth and length in Å.

	ΔG	ΔG with ZPE	dipole bowl	HOMO bowl	LUMO bowl	HOMO flat	LUMO flat	bowl depth low	bowl depth high	rim bond bowl	flank sub bowl	flank bond bowl	spoke bond bowl	Hub bond bowl	rim bond flat	flank sub flat	flank bond flat	Spoke bond flat	Hu bo fla
8 B3LYP/cc-pVDZ	7.85	6.99	1.24	-6.86	-2.79	-6.76	-2.90	0.50	0.83	1.40		1.44	1.39	1.42	1.41		1.45	1.37	1.
MP2/-//B3LYP/cc-pVDZ	8.46	7.60	2.15	-9.55	-0.14	-9.44	-0.27												
6 B3LYP/cc-pVDZ	9.56	9.21	0.25	-6.53	-2.26	-6.45	-2.38	0.53	0.88	1.39	1.44	1.44	1.39	1.42	1.40	1.46	1.46	1.37	1.
MP2-//B3LYP/cc-pVDZ	10.50	10.15	0.01	-8.88	0.71	-8.78	0.58												
7 B3LYP/cc-pVDZ	8.70	8.40	2.06	-5.70	-1.46	-5.63	-1.55	0.51	0.85	1.40	1.46	1.45	1.38	1.42	1.41	1.47	1.46	1.37	1.
MP2-//B3LYP/cc-pVDZ	10.71	8.34	2.41	-7.71	1.73	-7.62	1.65												
MP2/cc-pVDZ	10.78		2.41	-7.65	1.64	-7.55	1.53	0.54	0.90	1.41	1.46	1.45	1.40	1.42	1.43	1.47	1.46	1.38	1.
5 B3LYP/cc-pVDZ	9.03	9.38	1.05	-5.98	-1.77	-5.95	-1.90	0.52	0.85	1.40	1.46	1.45	1.39	1.42	1.41	1.47	1.46	1.37	1.
4 B3LYP/cc-pVDZ	11.53		3.23	-4.92	-0.92	-5.09	-1.00	0.51	0.84	1.40	1.46	1.45	1.38	1.42	1.41	1.47	1.46	1.37	1.
3 B3LYP/cc-pVDZ	24.33	22.82	7.13	-5.65	-1.61	-5.95	-1.90	0.52	0.86	1.40	1.46	1.45	1.39	1.42	1.41	1.48	1.46	1.37	1.

The molecular electrostatic map of several of these penta-substituted corannulene derivatives is presented bellow. The molecular electrostatic potential map of corannulene is presented in Figure 51.

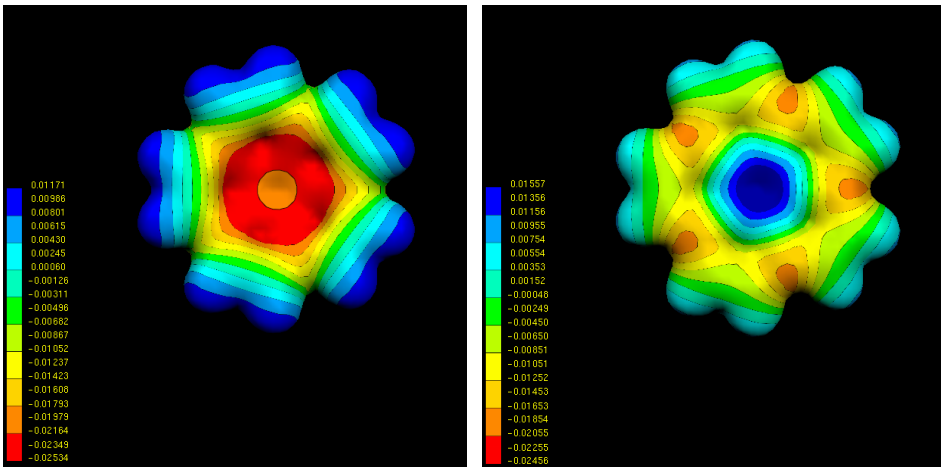


Figure 51: Molecular electrostatic potential map of corannulene, respectively bowl and flat (MP2/cc-pVDZ//B3LYP/cc-pVDZ).

We observe on Figure 51 that the electron density is more important at the bottom of the bucket formed by the bowl shape of corannulene (in red on the MEP), and the electropositivity presented in blue from the hydrogen atoms around the rim. This first molecular electrostatic potential will allow us to understand the changes in electrostatic potential when adding substituents to corannulene.

Figure 51 illustrates the molecular electrostatic map of pentamethylcorannulene, **7**, as referenced to corannulene. Here one observes that the electron density again concentrates at the center of the molecule in its bowl shape, with concentrated positive charge around the methyl groups. The flat form of the molecule shows a slightly different pattern in the electron density in the center of the molecule, where it is more diffuse. These map illustrate the effect of an electropositive substituent on corannulene.

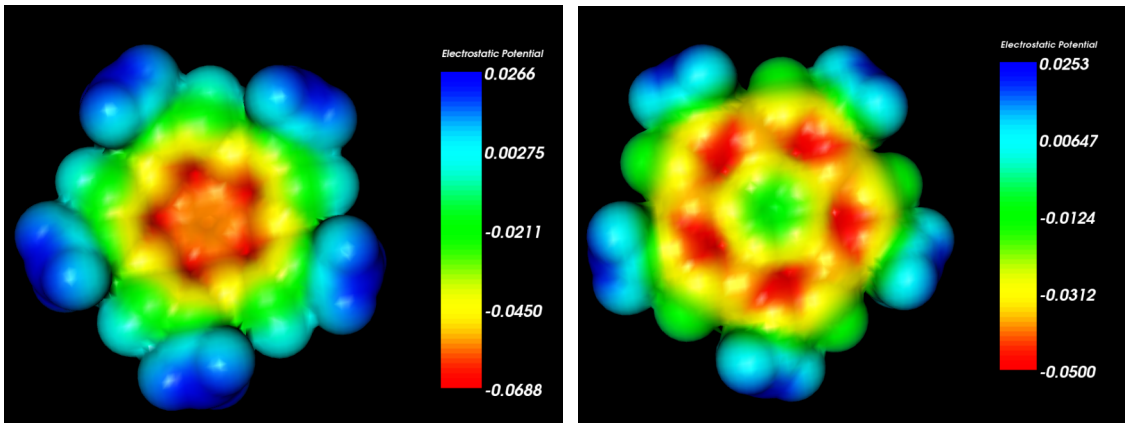


Figure 52: Molecular Electrostatic potential Map of pentamethylcorannulene respectively bowl and flat MP2/cc-pVDZ//B3LYP/cc-pVDZ.

Decafluorocorannulene, **8,** illustrates the effect of strong electronegative substitution does to the electronic distribution of corannulene. Analogous to benzene vs fluorobenzene, where the polarity of the molecule is reversed when fluorinated, so is corannulene when fluorine is added around the edge. For the bowl or flat form, one sees the negative character of the molecule localized around the rim of the bowl in the fluorinated form, as opposed to near the center in corannulene.

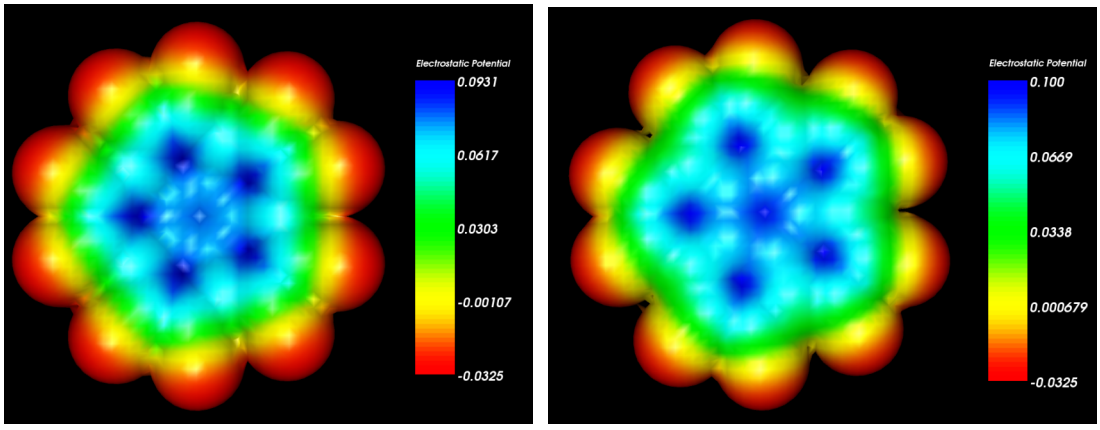


Figure 53: Molecular Electrostatic potential Map of decafluorocorannulene respectively bowl and flat MP2/cc-pVDZ//B3LYP/cc-pVDZ.

With only half as many fluorine substituents, pentafluorocorannulene, **6**, still shows a quite large influence of the fluorine on the electronic nature of corannulene, albeit attenuated from the above case.

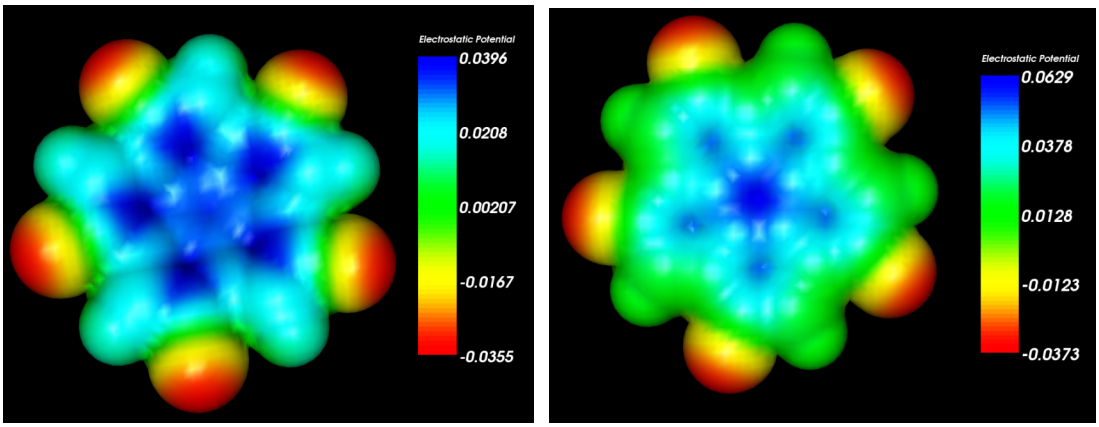


Figure 54: Molecular Electrostatic potential Map of pentafluorocorannulene respectively bowl and flat MP2/cc-pVDZ//B3LYP/cc-pVDZ.

It is interesting to observe the MEP for the pentadimethylphenyl-corannulene, **5**, given the discussion above concerning the unusually high barrier to interconversion and the impact of the methyl-methyl interactions. Here, we see the large positive nature of the adjacent methyl groups in the interior in the bowl conformation, with the phenyls are highly negatively charged, particularly near the attachment to the bowl structure. In the flat transition state conformation, these same methyl substituents are much less electropositive, and their respective phenyls also carrying more negative charge. One can imagine the difference in reactivity associated with these two conformational forms, depending on the solvent environment. For example, in the bowl structure, negatively charged solvent molecules of the commensurate size should interact very well inside the bowl structure, while in the transition state, preference would be for positively charge solvent molecules interacting with the flanking phenyl rings of the penta-substituents.

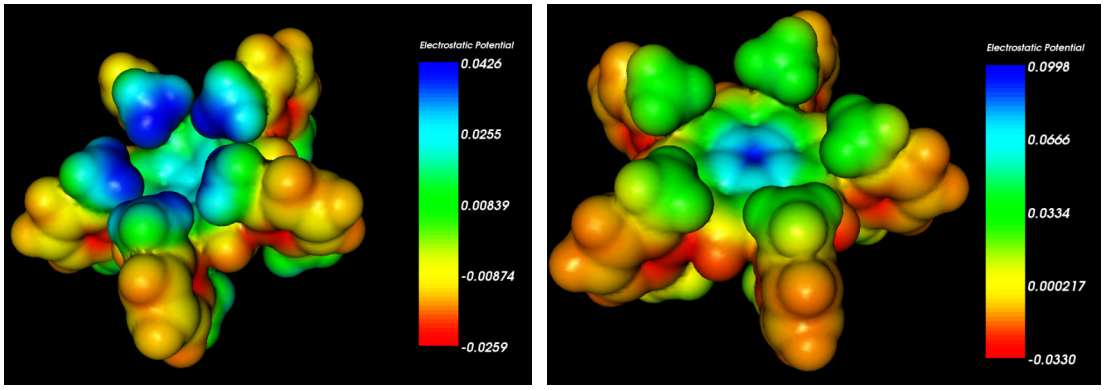


Figure 55: Molecular Electrostatic potential Map of penta(dimethyl phenyl)corannulene respectively bowl and flat B3LYP/cc-pVDZ.

The two oxygen atoms on both methoxy groups of the molecule **4**, and as the fluorine, the two atoms present the most electronegative character in the molecule. It is even more visible on the flat form of the molecule. As the molecule **5**, the electropositive behavior of the methyl group is more important in the flat form. The distance between two neighbors methyl group (hydrogen to hydrogen) is 4Å, which is compatible with stabilization from dispersion energy.

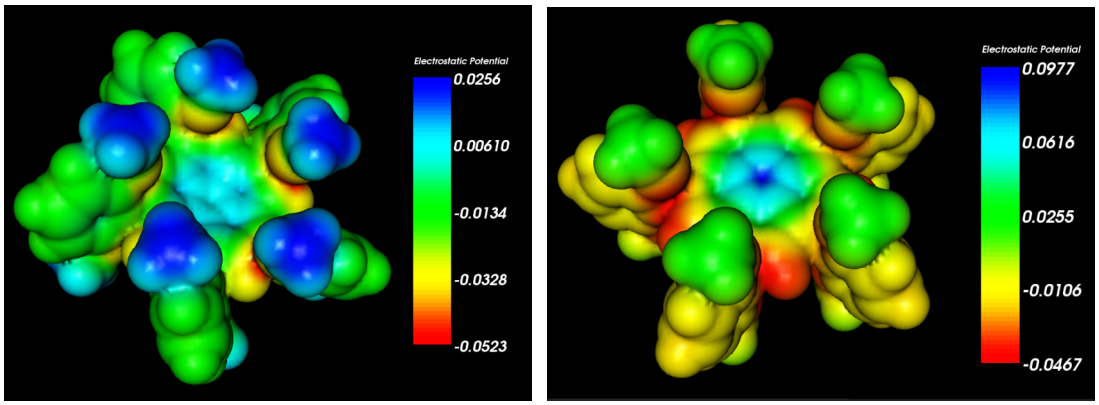


Figure 56: Molecular Electrostatic potential Map of penta(dimethoxyphenyl)corannulene respectively bowl and flat B3LYP/cc-pVDZ.

The last molecule of the set is the one with the lowest energy according to DFT calculations.

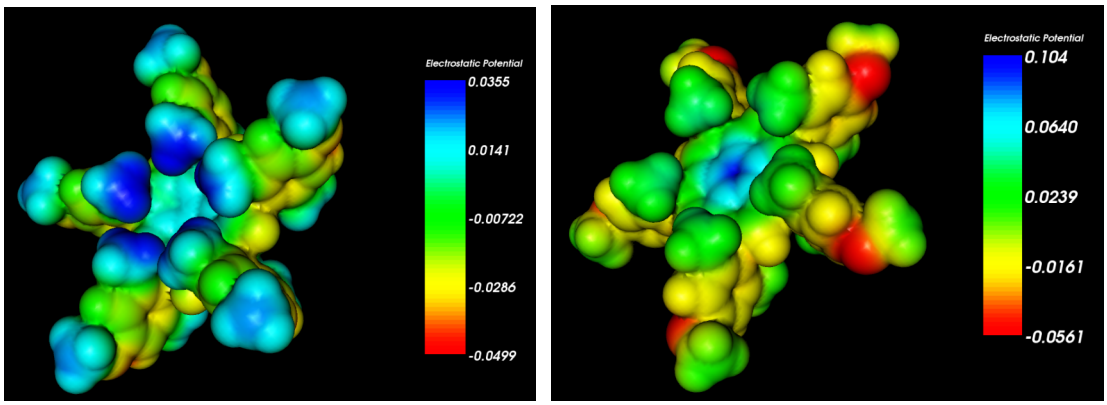


Figure 57: Molecular Electrostatic potential Map of penta(dimethylmethoxyphenyl) corannulene respectively bowl and flat B3LYP/cc-pVDZ.

Molecule **3** and **4** were studied in a solvent analysis since they were presenting unusually high barrier. This experimental study and partial theoretical results are presented in the next section.

5.3.3.3 Higher Order Nanostructures

Structure and properties of the series of PAH, $C_{20}H_{10}$, $C_{20}H_{30}$, $C_{40}H_{10}$, and $C_{50}H_{10}$, was previously undertaken by this group in 1997.[128] Results from that study were again reproduced and extended in the current work. A summary of results of the bowl depth and dipole values for each molecule is summarized in Table 22. Of particular interest here, is a focus on increasing carbon content to obtain greater bowl depth and observation of the changes in properties, such as dipole and charges distribution, from that of the basic corannulene molecule. On can see, for example, in Figure 58, that there is a linear increase in bowl depth with increase in carbon content, and a near perfect inverse quadratic relation of the dipole with increase in carbon content, where the dipole is observed to level off with increasing size of the bowl, to a little over 7 D at the largest PAH species considered. Given the information that the increasing bowl structures show convergence to a tube-structure, one might then imagine that the overall dipole of a bucky tube might be consistently around 7 D, but have linearly increasing bowl depth, dependent on the length of the tube structure.

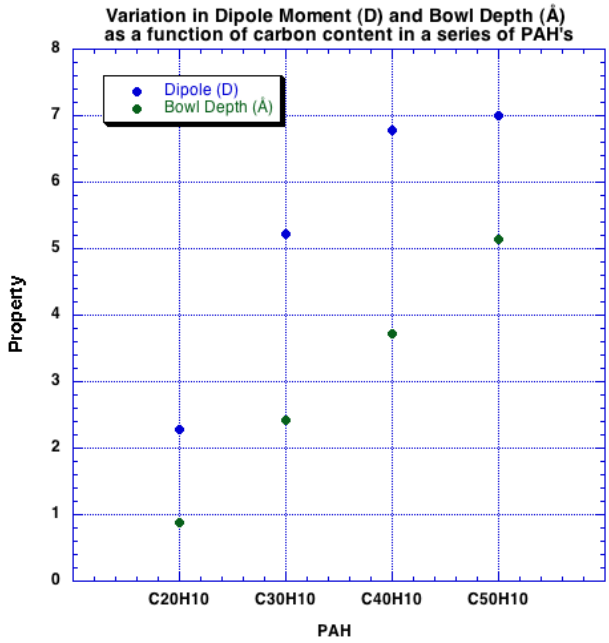


Figure 58: Variation in Dipole moment (D) and bowl depth (A) as a function of carbon content in a series of PAH's

Table 22. Dipole moment (Debye) and bowl depth (rim carbon) (Å) of the selected PAH, MP2/cc-pVDZ//B3LYP/cc-pVDZ.

Molecule	dipole	bowl depth
$C_{20}H_{10}$	2.29	0.89
$C_{30}H_{10}$	5.23	2.43
$C_{40}H_{10}$	6.79	3.73
$C_{50}H_{10}$	7.01	5.14

The molecular electrostatic potential MAP was determined for each of the PAH, to additionally illustrate how the increasing bowl depth from corannulene to the higher carbon content bowls influences charge distribution on the surfaces of these molecules.

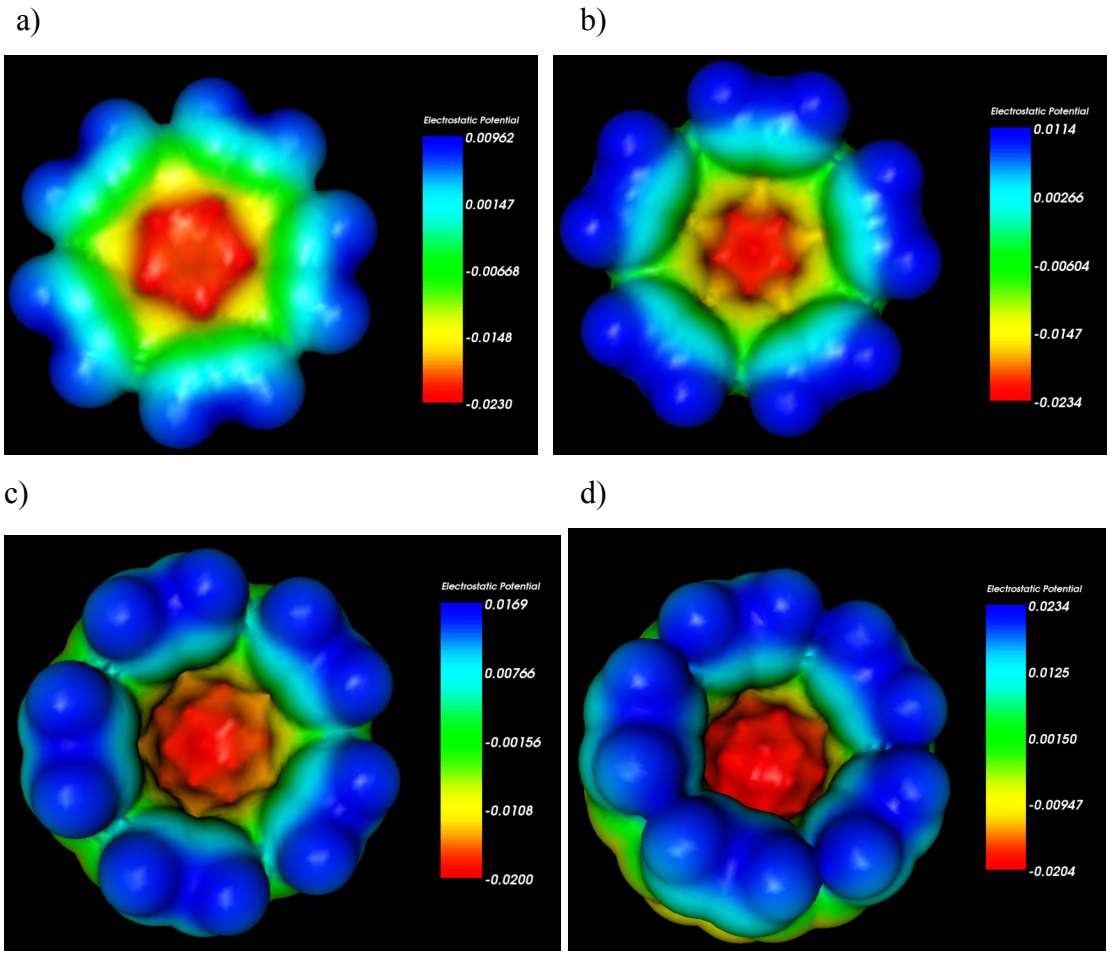


Figure 59: MEP map of PAH with increasing carbon content, using MP2/cc-pVDZ//B3LYP/cc-pVDZ. a) corannulene ($C_{20}H_{10}$) b) $C_{30}H_{10}$ c) $C_{40}H_{10}$ d) $C_{50}H_{10}$.

With the increasing bowl depth the electron density concentrates more and more at the bottom of the bowl, resulting in the increase of the dipole. These results enable better understanding of the potential for reactivity of these surfaces with other reactant species, for example, their potential for complexation with metals.[140] The following studies are also focused on the modification of the bowl depth and geometrical properties with the addition of substituents.

5.3.3.4 Penta-substituted corannulene

A particularly interesting substituted corannulene is the previously synthesized penta-corannulene corannulene, …. [The other paragraph was for the penta-dimethyl-aryl-corannulene, not the penta-corannulene corannulene, so, you need to just write this paragraph according to the data that you produced and are showing in the Table below. You need to compare to, for example, corannulene itself, and the

other penta- systems with respect to 1) paragraph on structure, 2) paragraph on bowl depth/inversion barrier, and then on dipole and predicted solvation)

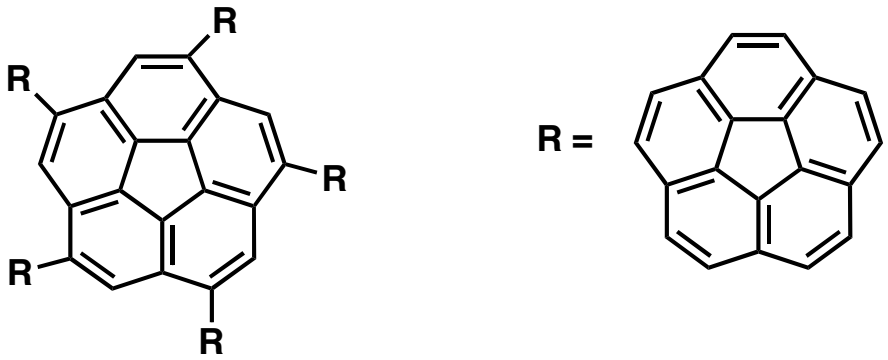
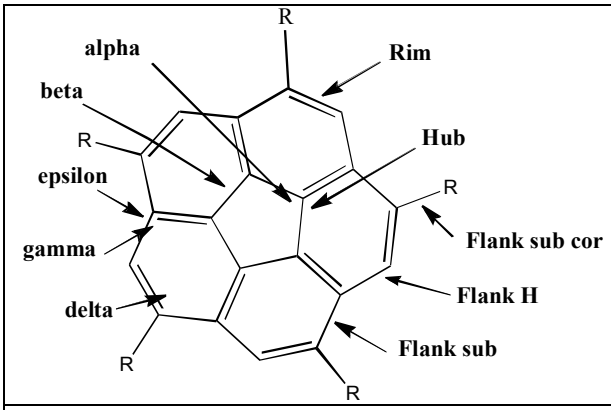


Figure 60: structure of *sym*-pentacorannulylcorannulene.

Table 23: Structural information and properties of *sym*-pentacorannulylcorannulene (bond lengths in Å, and angles in degrees).

	**B3LYP/cc-pVDZ**	**B97-D/cc-pVDZ**
hub bond	1.4162	1.4222
spoke bond	1.3828	1.3942
rim bond	1.4016	1.4082
flank sub	1.4645	1.4657
flank H	1.4516	1.4509
bond flank sub Cor	1.496	1.4906
bowl depth spoke	0.4794	0.5309
bowl depth flank H	0.7816	0.8779
bowl depth flank sub	0.7892	0.8863
distance hub hub	9.4056	9.3996
distance H-H sub 1-3	22.3482	21.6682
POAV angle	97.48	98.25
alpha angle	108.0	108.0
beta angle substituent side	123.82	123.24
beta angle H side	123.09	122.61
delta angle substituent	120.56	120.42
delta angle H	123.17	122.85
epsilon angle	131.25	129.48
gamma angle substituent	114.14	114.8
gamma angle H	113.78	114.55

dipole (D)	3.48	4.82
ionization potential (HOMO) eV	-5.91	-5.19
symmetry HOMO	E2	E1
dispersion energy (kcal/mol)	0	-179.12

The molecular electrostatic potential maps (MEPs) for both B3LYP and B97-D predicted structures were also carried out. The scale of electrostatic potential was taken the same to illustrate the relative variance with and without dispersion capabilities, (shown in Figures 61-62).

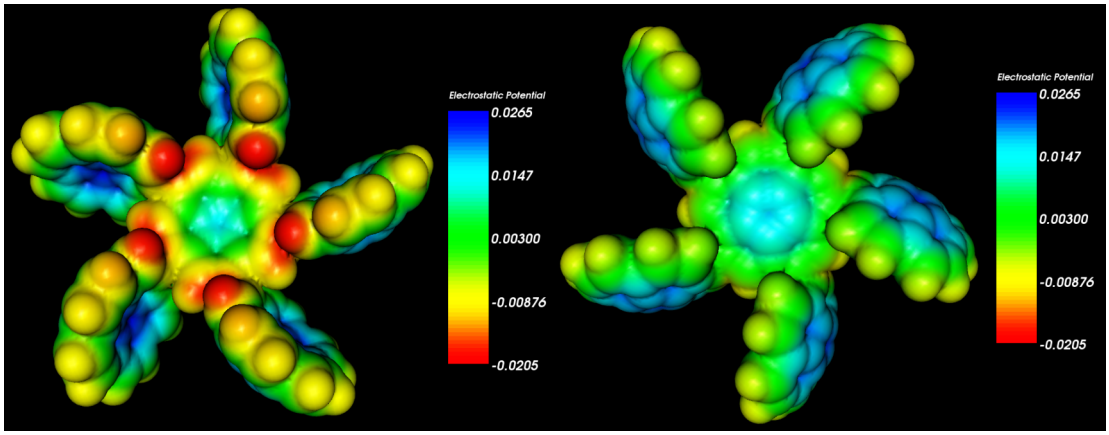


Figure 61. B3LYP/cc-pVDZ MEP map of compound 11, view of concav side (left), and b) view of convex side (right).

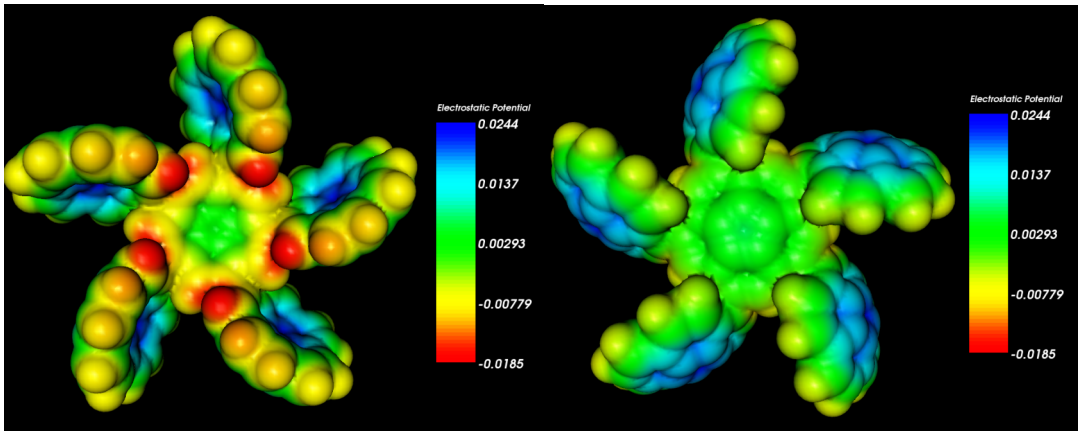


Figure 62. B97-D/cc-pVDZ MEP map of compound 11, view of concav side (left), and b) view of convex side (right).

Comparing these two sets of MEPs, one can observe the effects of accounting for dispersion interactions in the method. In the B3LYP method, without explicit

accounting for dispersion, one sees that the base of the convex side of the bowl is less electrophilic than it is the more representative B97-D method. One sees the similar effect on the convex side of the bowl interior. In terms of interactions with surrounding media, the interior of the bowl could potentially interact with a solvent molecule, as could the convex side of the bowl. As such, the ability to properly account for these would benefit from the higher order method.

The difference in predicted dipole between B97-D (4.81) and B3LYP (3.47) also is indicative of the ability to handle dispersion to properly model this property. The dipole also provides an indication of expected solvent effects, suggesting that the solvent predictions would be greatly enhanced using the B97-D method. Moreover, these dispersion enabled DFT methods should also be used with the new methods of inclusion of non-electrostatic effects that have been presented, i.e., DCOSab-RS + B97D.

5.4 Perturbation / Influence of the Environment

5.4.1 Solvent Effect on Inversion Energies

As already mentioned, despite the large number of derivatives used to establish the strong structure-energy correlation of bowl depth with bowl-to-bowl inter-conversion barrier, exceptions have been found. Hayama[139] reported an inversion barrier of 12.1 kcal/mol for **3** (Figure 64), which is 2.5 kcal/mol higher than would be predicted by the structure/energy correlation function. Density functional theory computations using B3LYP/cc-pVDZ predict an inversion barrier of 8.5 kcal/mol, while the generally more accurate MP2/cc-pVDZ//B3LYP/cc-pVDZ method predicts a barrier of 14.6 kcal/mol.[141] It was suggested that the unusual high inversion barrier arises from van der Waals attractive interactions among the *endo* methyl groups of the phenyl substituents in the bowl conformation of **3** (Figure 63), resulting in an overall higher barrier of inversion.

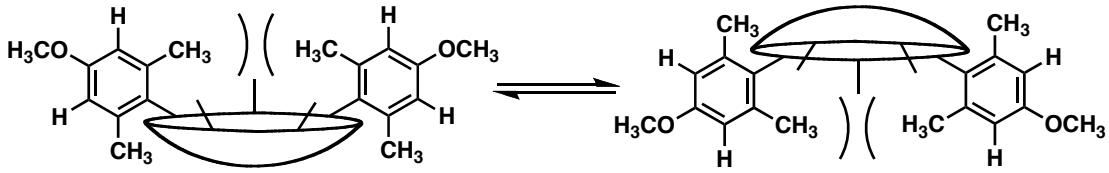


Figure 63: van der Waals attractive forces among the *endo* methyl groups.

Experimental and theoretical studies of the correlation between the isotopic effect and ΔG_{inv} was also presented,[141] again suggesting that the van der Waals interactions of $CD_3//CD_3$ are stronger than these of $CH_3//CH_3$, and providing further evidence of the increased stability in the bowl structure relative to the transition state structure, resulting in a larger barrier to interconversion.

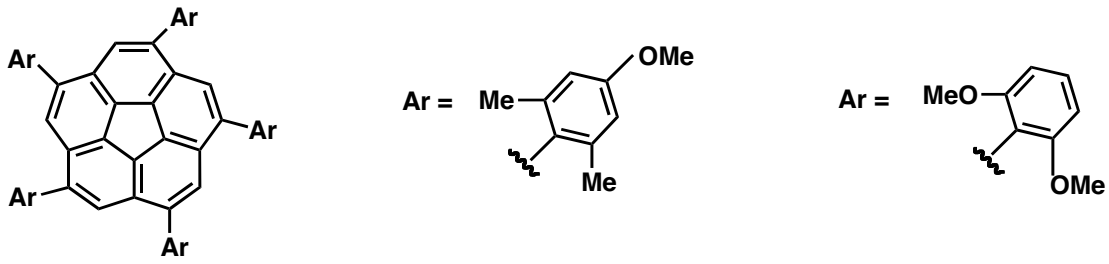


Figure 64: Corannulene derivatives with unusual inversion barrier: penta(dimethyl methoxy)corannulene (3) and penta(dimethoxy)corannulene (4).

We saw in the previous section that functionalization of corannulene affects not only the structural properties of the molecules but also its properties, such as dipole. An explicit interaction of individual solvent molecules could also be responsible for an increase in inversion barrier, if the interaction again preferentially is stabilized in the bowl structure relative to the transition state structure.

From experimental results, it appears that solvent can play a role in the dynamics in several derivatives of corannulene. The role of the solvent are thought to result from two different types of influence dependent on the specific functionalization on the corannulene nucleus. Specifically, in one derivative, the barrier to interconversion was seen to be effected by solvent volume, while another derivative was seen to be effected by solvent dielectric. The results from experiment is shown in Table 24.[141]

Table 24: Experimental ΔG using line-shape analysis of several solvents for 3 and 4. Value in brackets is derived from the mean plane of the hub carbons to the mean plane of the spoke atoms.

solvent	ΔG_{inv} kcal/mol	
	3	4
Toluene	10.48 (0.027)	10.63 (0.053)
Chlorobenzene	10.33 (0.007)	-
Chloroform	11.00 (0.013)	10.70 (0.032)
Dibromomethane	11.89 (0.025)	-
Dichloromethane	11.73 (0.017)	11.50 (0.018)
Acetone	11.07 (0.17)	12.41 (0.020)
DCFM [b]	11.20 (0.029)	-
Bowl Depth (Å) [a]	0.94 [0.56]	0.91 [0.54]

[a] **dichlorofluoromethane**

[b] **DCFM**

As mentioned in chapter 2, there are a variety of effects that can result from interaction with solvents. While these typically relate to polarity, the physical properties of some of the solvents can also be important; that is, the larger the solvent molecule, the greater dispersion forces can be, disregarding steric effects. A general break down of properties as a function of solvent are collected in Table 25.

Table 25: Some physical properties of solvent.

solvent	M_T [a]	ρ [b] (g/cm^3)	V [c] (Å^3)	ε_r [d]	α [e] 10^{-24}(cm^3)
Toluene	92.14	0.8669	176	2.379	12.26, 11.8
Chlorobenzene	112.56	1.1058	169	5.6895	14.1, 12.3
Chloroform	119.38	1.4832	134	4.8069	9.5, 8.23
Dibromomethane	173.83	2.4969	116	7.77	9.32, 8.62
Dichloromethane	84.93	1.3266	106	8.93	7.93, 6.48
Acetone	58.08	0.7899	122	21.01	6.4, 6.39, 6.33
DCFM	102.92	1.405	122	—	9.5, 6.82

[a] M_T : molecular weight. [b] ρ : density at 20°C (9°C for DCFM).[15] [c] V : volume, $V = \frac{M_T}{\rho} \times \frac{1}{N_A} \times 10^{24}$.

[d] ε_r : dielectric constant.[15] [e] α : polarizability.[15]

The difference in bowl-to-bowl interconversion barrier for **3** and **4** is suggestive of a solvent effect of some type. The above discussion involving the dispersion effects due to the interactions of *endo* groups in the bowl conformation, are

expected to be solvent dependent in some way. In the case of **3**, a correlation between ΔG_{inv} and solvent volume has been suggested, with smaller solvent volume resulting in larger ΔG_{inv}. In the case of **4**, a correlation between ΔG_{inv} and solvent polarity, defined by the dielectric constant ε, has been suggested. Since **4** has a dipole of 3.2 D in ground state (bowl) and 0 D in transition state (flat), it is expected that polar solvents will stabilize the ground state preferentially, resulting in an increase in the barrier. One might also expect further stabilization with consideration of an explicit solvent molecule in the interior of the bowl, depending on the actual solvent considered. Stabilization by solvent molecules could additionally lead to increase order in the solvent cage, and smaller vibrational movement in the ground state.

5.4.1.1 Theoretical Work

The structure and properties of corannulene as a function of organic solvent has been carried out, the results of which are presented in Table 26.

Table 26: Energetics and Properties of corannulene in solvent.

B3LYP/cc-pVDZ dblcav	acetone	benzene	CCl_4	chloro-benzene	chloroform	DCFM	dibromo-methane	dichloro-ethane	dichloro-methane	methanol	THF	toluene	water
Bowl													
gas phase energy	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73
Total solvated Energy	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73
Total Energy corrected	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73	-767.73
OCE kcal/mol	0.15	0.08	0.08	0.13	0.13	0.13	0.14	0.15	0.14	0.16	0.14	0.09	0.16
Solvated Energy	1.85	-2.54	-2.52	1.82	1.80	1.81	1.85	1.86	1.86	1.83	1.85	-2.68	1.81
Dielectric Energy	-0.0095	-0.0038	-0.0037	-0.0071	-0.0067	-0.0069	-0.0079	-0.0085	-0.0082	-0.0099	-0.0079	-0.0040	-0.0103
Sum of polarization charges	0.0316	0.0323	0.0323	0.0319	0.0319	0.0319	0.0318	0.0317	0.0317	0.0316	0.0318	0.0323	0.0315
surface area of cavity (A**2)	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55	917.55
volume of cavity (A**3)	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47	1976.47
Flat													
gas phase energy	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71	-767.71
Total solvated Energy	-767.71	-767.72	-767.72	-767.72	-767.72	-767.72	-767.72	-767.71	-767.72	-767.71	-767.72	-767.72	-767.71
Total Energy corrected	-767.71	-767.72	-767.72	-767.72	-767.72	-767.72	-767.72	-767.71	-767.72	-767.71	-767.72	-767.72	-767.71
OCE kcal/mol	0.14	0.08	0.08	0.12	0.11	0.12	0.13	0.13	0.13	0.14	0.13	0.08	0.14
Solvated Energy	2.98	-2.58	-2.56	-4.44	-4.22	-4.36	-4.89	3.22	-5.06	2.85	-4.86	-2.72	2.69
Dielectric Energy	-0.01	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.01
Sum of polarization charges	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
surface area of cavity (A**2)	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83	924.83
volume of cavity (A**3)	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50	1980.50
dielectric constant	20.70	2.25	2.23	5.62	4.90	5.34	7.77	10.36	8.93	32.63	7.58	2.38	78.39
ΔE barrier	10.17	9.00	9.01	2.79	3.03	2.88	2.30	10.40	2.13	10.05	2.34	9.00	9.92
volume	584.44	743.18	866.45	897.98	714.34	614.00	644.51	706.11	558.17	328.19	671.83	893.40	172.73
ΔE barrier without OCE	10.18	9.01	9.01	2.80	3.04	2.89	2.32	10.41	2.14	10.07	2.35	9.01	9.94

A visual display of the correlation between the dielectric constant and bowl-to-bowl inter-inversion barrier is illustrated in Figure 65. One can indeed see a solvent effect, but only by a total of 0.5 kcal/mol in either case, so effectively a very small change due to effects of solvent. This also agrees with experimental findings.

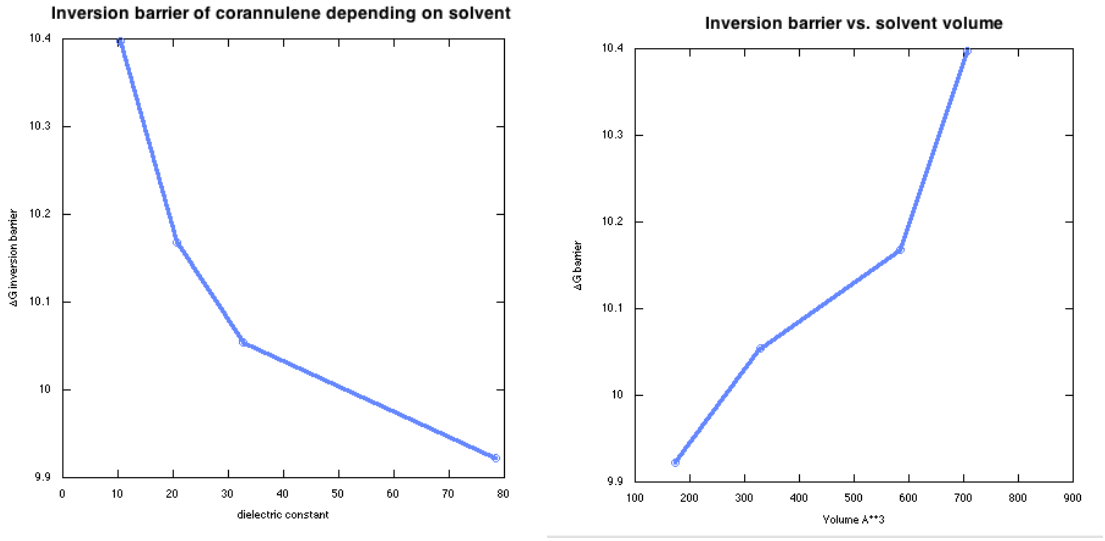


Figure 65: Inversion barrier (kcal/mol) as a function of dielectric constant of the solvent (left), and as a function of solvent volume (right).

Since the structural and properties of the solvent can influence the dynamics of the corannulene, which as previously observed in the case of the benzene isomers can be masked by simple dipole and electrostatic energy arguments, it is interesting to see results in terms of the σ-profiles discussed earlier. For comparison, the σ-profile of benzene is shown as a reference with corannulene (Figure 66).

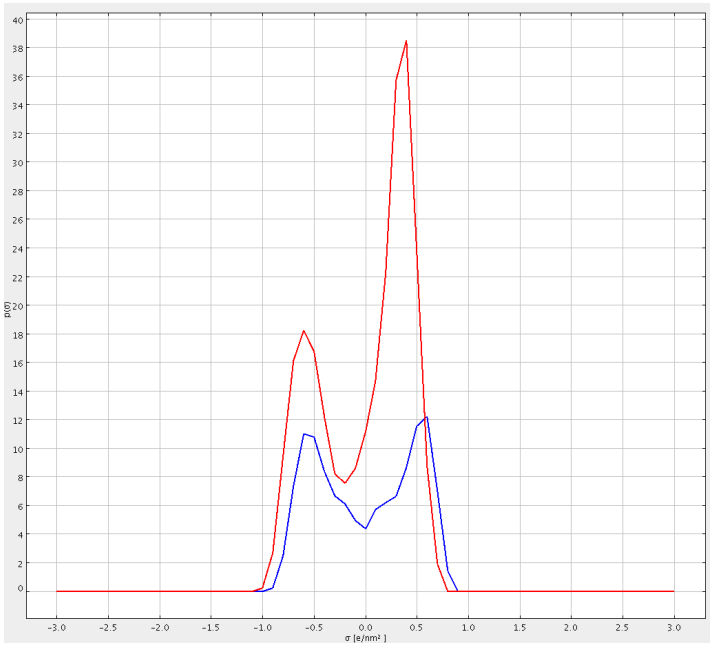


Figure 66: sigma profile of corannulene (red) and benzene (blue): BP86/TZVP in water.

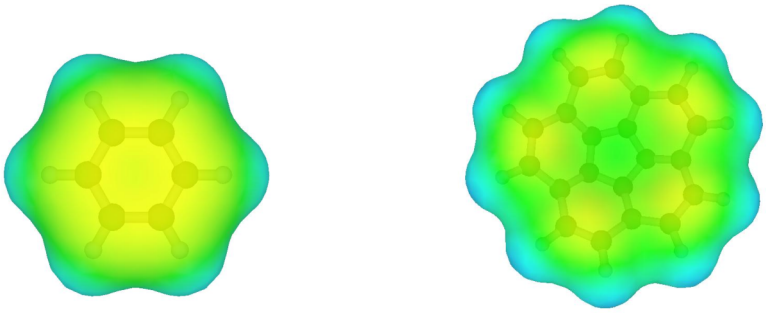


Figure 67: sigma surfaces of benzene (left) and corannulene (right) from COSMOtherm, BP86/TZVP in water.

What is interesting is that the two structures appear to be structurally quite similar according to where their peaks in terms of sigma units sit. The pictorial representations of the surfaces also show this, with their belt regions showing delta positive, and their interior ring regions showing more delta negative, albeit small (as clear by the nearness of the sigma values to zero). The increased surface area of the corannulene relative to the benzene is emphasized by the height of the peaks compared to the same relative peak in benzene. This analysis would suggest that corannulene and benzene should show similar reactive characteristics with other molecules, albeit a significantly larger surface area from which to interact with in the former molecule.

The other functionalized corannulenes described here, **3** and **4**, while of great interest to have verifying solvation studies, was discovered not to be possible at this time due to the enormous size of the cavity, which is still limited. Subsequent to the finishing of this PhD thesis work, however, this capability has now been enabled by another PhD student, and it is expected that these systems will now be carried out to determine the comparisons to experiment of solvent volume vs. solvent dielectric effects.

5.4.2 Crystal Packing

The molecular structure of corannulene has been determined by X-ray crystallography, which enabled to evaluate the value of the bowl depth (0.875Å).[123,142] Several derivatives have also been characterized after synthesis.[143] Three substituted corannulene are studied in the section, even though they present the same number of substituent and position, the crystal packing of the three molecules is different. The electronic structure of each molecule is studied and the crystal package presented in chapter 4 is used here to explain what are the forces influencing the observed packing.

Among all the possible structures, we focus here on three tetra-substituted derivatives as shown in Figure 68.

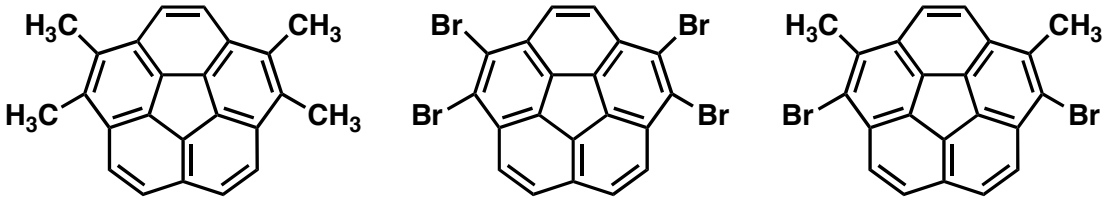


Figure 68: Tetra substituted derivatives of corannulene. 1,2,5,6-tetramethylcorannulene (9); 1,2,5,6-tetrabromocorannulene (10) and 1,6-dibromo-2,5-dimethylcorannulene (11).

The interesting feature of these three derivatives is that they belong to different crystallographic space group, even though their volume is comparable, with **9** (Figure 68) belonging to the $P2_1/n$ space group, **10** (Figure 68) to the $Pna2_1$ space group, and **11** (Figure 68) to the P21/c, the same experimentally determined space group as corannulene (**1**).

$P2_1/c$ space group is the most common space group for organic molecules. This space group can lead to different notation according to the conventions adopted. In $P2_1/c$, $P2_1/a$ and $P2_1/n$, the unique axis is *b* (screw direction), but in the first case ($P2_1/c$) the glide plane direction is oriented along *c*, in the second it is along *a*, and in the third it is along the *ac* diagonal. The particular notation is sometimes dictated by the need to keep the angle β as close as possible to 90˚. When different settings are used, it becomes necessary to specify in detail which axis is the unique one with a complicated conventional notation. For example, the full notation for molecule **11** is actually P 1 21/c 1, meaning that the main screw axis is along *b* and is perpendicular to *c*.

5.4.2.1 Theoretical Work

Professor Gavezzotti has previously illustrated the use of the *Prom* module of the PIXEL software (described in chapter 4) on several derivatives of corannulene (Cl-,dicarbonyl-, and a carboxylic acid derivative) for possible polymorphic structure, in his book[4] entitled "Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids". Following the same idea, we are interested in looking at polymorphic structure in several derivatives, and further discussing the differences in packing characteristics. In particular, the three derivatives of corranulene of interest are tetrabromo-, tetramethyl-, and dibromo,dimethyl-coarnnulene. To carry this analysis out, we make use of the Opix module of the PIXEL software, to investigate structures and properties. A diagram explaining the different modules of the Opix software is available in chapter 4.

The crystallographers provided the CIF structural files for **9**, **10**, and **11**.[144] Reproductions of the crystal structure made from the computational software for the three systems are shown in Figure 69,70 and 71.

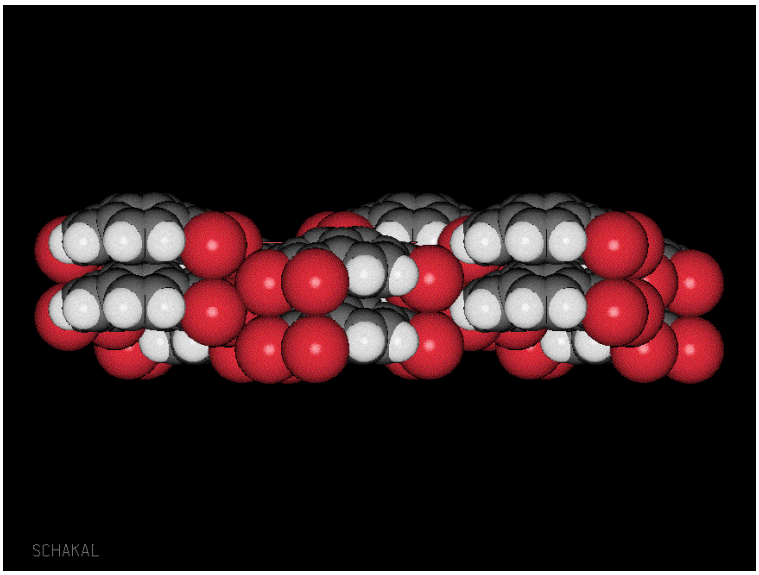


Figure 69: Representation of crystal of 1,2,5,6 tetrabromocorannulene: Pna 21 crystal packing.

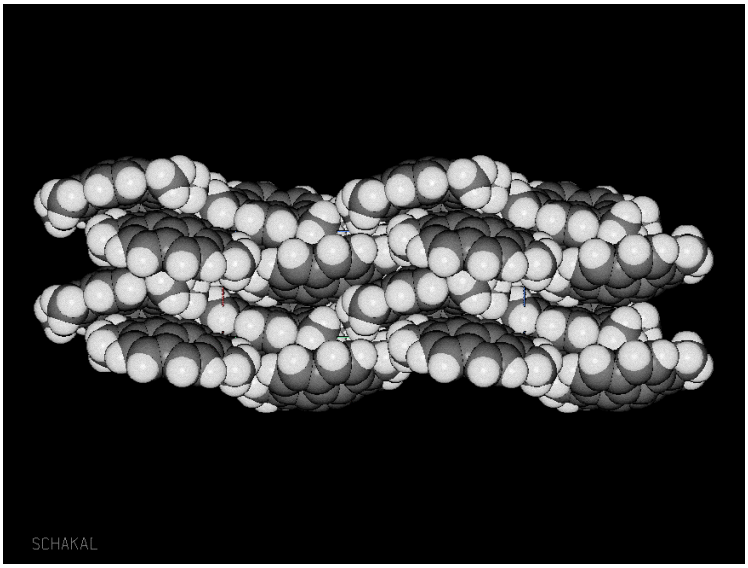


Figure 70: Representation of crystal of 1,2,5,6 tetramethylcorannulene: P21/n crystal packing.

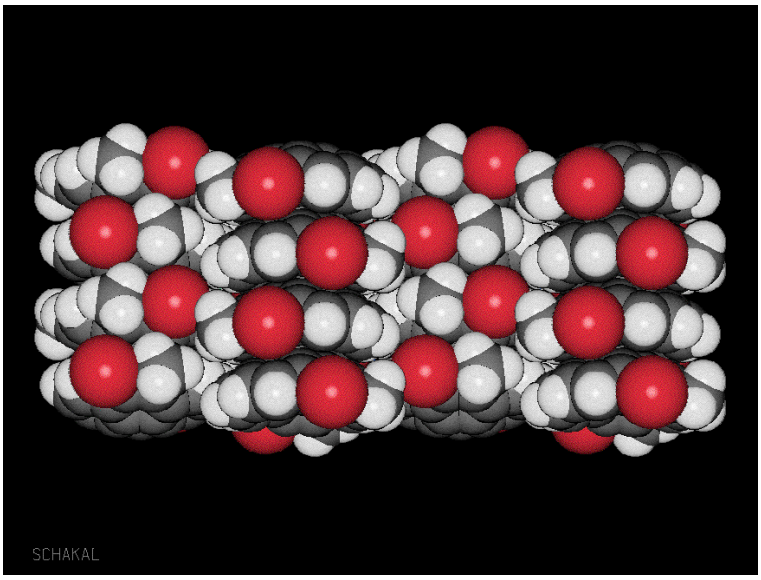
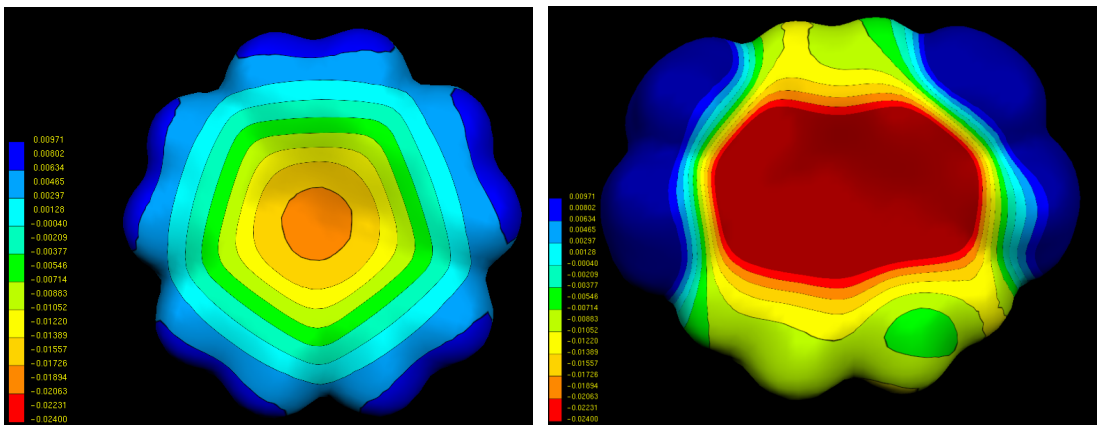


Figure 71: Representation of crystal of 1,6 dibromo-2,5 dimethylcorannulene: P21/c crystal packing.

Each molecule is observed to pack in a different space groups, and from Figures 69, 70 and 71, one can see that the organization of the crystal is clearly different among the three. The present chapter shows a series of calculations on these three molecules in order to understand the reason of such differences.

Starting from the experimental X-ray crystal structure, corresponding molecular electrostatic potential maps were calculated for each species in order to better understand the electronic structure and properties at that structure. In order to have a view of the electrostatic of the corannulene itself, the unsubstituted corannulene is also shown in Figure 72a on the same scale.

a) b)



c) d)

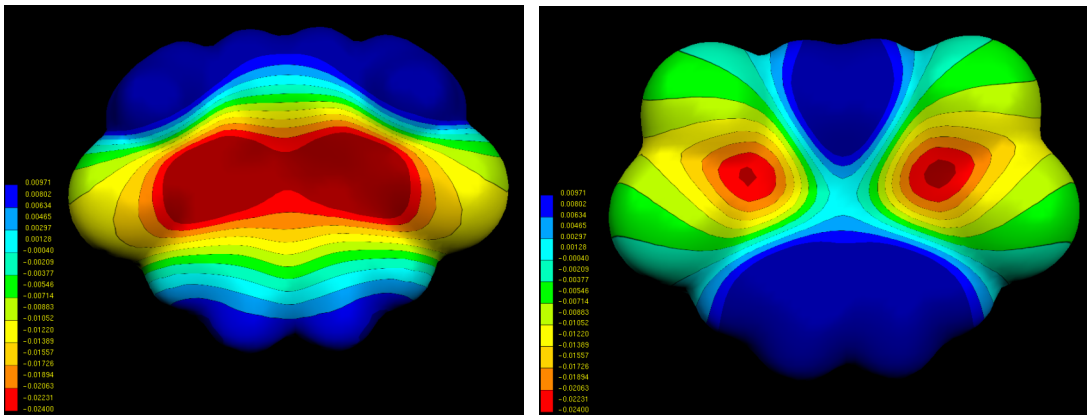


Figure 72. MP2/cc-pVDZ calculated molecular electrostatic potential maps of, a) corannulene, b) 1,2,5,6-tetramethylcorannulene, **9**, c) 1,6-dibromo-2,5-dimethylcorannulene, **11**, d) 1,2,5,6-tetrabromocorannulene, **10**.

The electron density rich area is concentrated in the five-membered ring in the case of the unsubstituted corannulene (Figure 72a). The hydrogen surrounding the rim presents the electropositive belt of the hydrogens. Since all the molecular electrostatic potentials are presented with the same scale, one can compare the donating and withdrawing properties of each molecule depending on the methyl and bromo substituents. With four methyl group donating substituents in **9**, one sees a focusing of negative charge in the center of the molecules, showing a broad capacity to act as an electrophile in this region, while the methyl groups themselves are quite electropositive. In contrast, when all four are bromine atoms as in **10**, the electron withdrawing capability of the bromines draws out much of the electron density of the corannulene ring, and there is only small focus of electronegative area near the base of the bromine substituents, and a much broader area of electropositive nature. Finally, in between these two extremes, with two electron withdrawing and 2 electron donating substituents, **11**, one sees something in between the other two MEP maps, with a significant but focused electron rich area in the center of the molecule, and a broad region of electropositive nature across the top and bottom. These electrostatic maps can help explain reactivity of the different molecules and how they may pack in type A, B or C motifs.

To better understand the actual differences in packing characteristics in these three systems, we can investigate the lattice energy of each system as represented in the space group of its experimentally known x-ray crystal structure. This is carried out using the program OPROP, described in 4.2.4. The results for calculated volume, lattice energy, and non-bonded interaction energy, for each of the three systems, are summarized in Table 27.

Table 27: Crystal packing properties for **9, 10,** and **11**, calculated using the OPROP module, at the MP2/cc-pVDZ//B3LYP/cc-pVDZ level of theory.

Molecule	Volume (A^3)	Lattice Energy (kJ/mol)	Non-bonded interaction Energy (kJ/mol)	Space group
9	281.1	-159.4	-63.8	P21/n
10	290.6	-170.2	-103.4	P21/c
11	299.2	-185.9	-104.7	Pna21

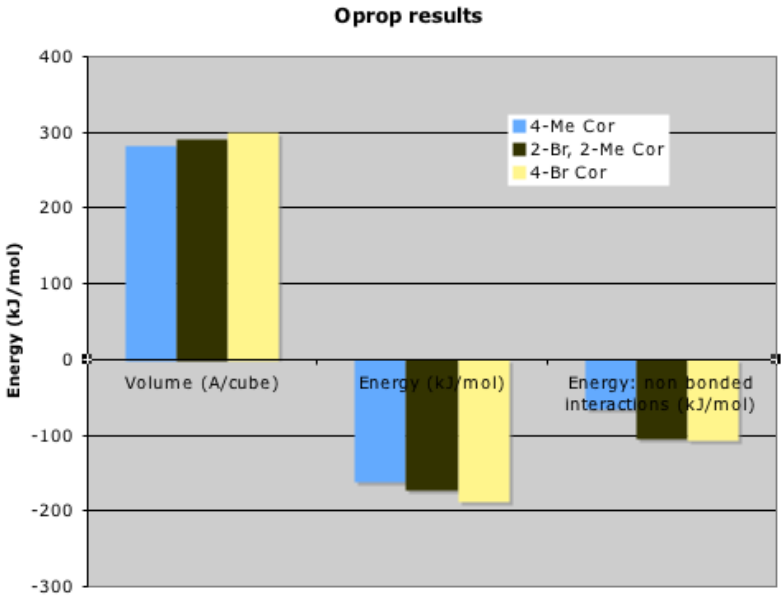


Figure 73. Crystal packing properties for **9, 10,** and **11**, calculated using the OPROP module, at the MP2/cc-pVDZ//B3LYP/cc-pVDZ level of theory.

In comparing the three systems, **9**, **10**, and **11**, one sees a very small, but gradual increase in volume with replacement of methyl substituent with bromine. This is accompanied by a gradual increase in lattice energy with increasing number of bromine substituents. Even more dramatic is the difference in non-bonded interaction

energy when comparing the tetra-methyl substituted case with either of the bromine substituted systems (Table 27 and Figure 73).

In a recent related experimental paper of our collaborators,[145] it was proposed that the molecule packing in the crystal depends on the size of the bowl surface area and on the depth of the bowl. A larger and deeper bowl was suggested to favor a columnar packing of the molecules, with bowls of neighboring columns either oriented in the same direction (resulting in a polar crystal) or in the opposite direction (resulting in an apolar crystal). As the size of the surface area of the molecule decreases, the molecular packing would tend to arrangements different from a columnar structure. This scheme has been confirmed by a variety of studies on substituted corannulene,[140] where unidirectional columnar molecular packing was observed, either in orthorhombic or monoclinic unit cell.

In the present study, tetrabromocorannulene, **10**, has the deepest bowl (0.90Å), in comparison to dibromo,dimethylcorannulene, **11**, or tetra-methylcorannulene, **9**, (085Å and 0.81Å, respectively). Additionally, the surface area of **10** is the largest, due to the size of the bromine atoms. The tetrabromo- derivative is also the only crystal structure associated with an orthorhombic space group, Pna21, with a polar column pattern. In contrast, the di-bromo,di-methyl derivative shows an apolar stacking within the space group $P2_1/c$. The patterns are clearly visible in Figure 69, 70, 71. The tetramethyl- derivative, having a shallow bowl depth, does not show a columnar arrangement in the crystal environment, while its experimental space group is $P2_1/n$.

To further understand the various packing forms of these three different functionalized corannulene structures, it is useful to look at the possible polymorphic structures for each case. Using the PROM module described above, the theoretical polymorph predictions for each of these three structures can be made. One has to interpret these results with some caution, however, since the low energy structures generated by this theory will not always corresponds to that experimentally observed. There are several additional factors that also must be considered in the interpretation, for example, differences in experimental conditions, effects of solvent, etc.[34]

The idea is to consider packing of each of the three functionalized corannulene structures into all three space groups, in order to determine the potential accessibility of these forms to the alternative packing of the related conformer. Unfortunately, a

couple of simplifications had to be accommodated for due to the limitations in space group choices within the Prom module. For example, tetramethylcorannulene adopts the $P2_1/n$ symmetry, but the P21/c symmetry was chosen since the two space groups are only different in that the P21/c group has only an additional halftorsion along the *x* axis. Additionally, the module did not offer $Pna2_1$, but the related $P2_12_12_1$orthorhombic crystal form was available. The space group $P2_1$ was also investigated as a monoclinic crystal arrangement. Packing results for 1,2,5,6-tetramethylcorannulene, **9**, 1,6-dibromo-2,5-dimethylcorannulene, **11**, and 1,2,5,6-tetrabromocorannulene, **10** are show in Tables 28, 29, and 30, respectively. In these tables, energetics are broken down into the following contributions: coulombic, dispersion, repulsion and polarization. The corresponding packing arrangements as predicted by the PROM module are shown in Figures 74, 75 and 76.

1,2,5,6-tetramethylcorannulene is observed experimentally to adopt a $P2_1/n$ packing arrangement. Analyzing the theoretical results in terms of the total of the coulomb, dispersion, repulsion and polarization contributions, one sees that the $P2_12_12_1$ is the favored form, although only ~10 kJ/mol lower in energy that the other two forms. Comparing the various individual contributions, one finds that the dispersion contribution, the main stabilization energy in molecular crystals, has also the greatest contribution in the $P2_12_12_1$ form, and the lowest repulsive energy. P21/c and $P2_1$ appear to be similar in all energy contributions except the Coulomb energy, which in the $P2_1/c$ form is the lowest of all three packing forms. The $P2_1/c$ form also has a noticeably higher number of generated polymorphs than either of the other two forms (23 vs 9 or 4), but it also has the largest range of energies for these forms (~40 kJ/mol range, compared to ~20 and ~10 kJ/mol range for the other two). Figure 74 shows the represented crystal packing structural forms for the three space groups.

Table 28. Energetic contributions for the different space group arrangements generated from the PROM and PIXEL modules (energy in kJ/mol) for 1,2,5,6-tetramethylcorannulene.

Energetic Contribution	$P2_1$	P21/c	$P2_12_12_1$
coulomb	-34.1	-40.7	-35.5
dispersion	-229.7	-228.3	-234.5
repulsion	148.9	152.2	141.1
polarization	-35.1	-35.6	-33
Total	-150	-152.4	-161.9

(coulomb+dispersion+ repulsion+polarization)			
6-exp (equation 37)	-164.3	-163.9	-167.1
# of polymorphs generated	9	23	4
Range of energy of predicted polymorphs	-144.4,-164.3	-125.2,-163.9	-157.2,-167.1

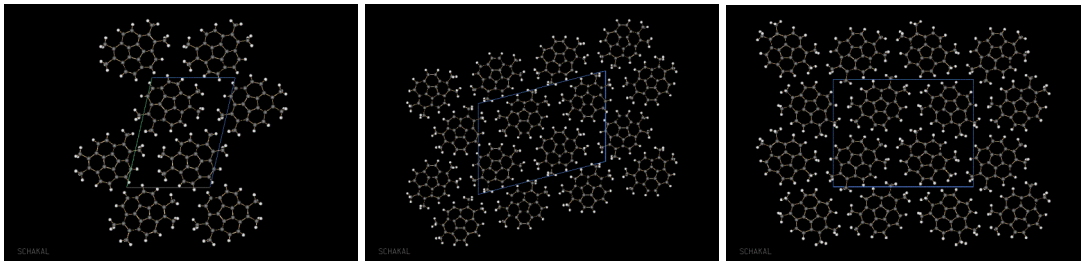


Figure 74. Crystal structures of 1,2,5,6-tetramethylcorannulene as predicted by the PROM module, and used with the PIXEL analysis. (a) $P2_1$ space group. (b) P21/c space group. (c) $P2_12_12_1$ space group.

1,6-dibromo-2,5-dimethylcorannulene is observed experimentally to adopt a $P2_1/c$ packing arrangement. By looking at the theoretical results presented in Table 29 in term of the total of the coulomb, dispersion, repulsion and polarization contributions, one sees that the $P2_12_12_1$ is the preferred form, it is ~10 kJ/mol lower in energy than the $P2_1$ form and ~15 kJ/mol lower than $P2_1/c$. Comparing the various individual contributions, the dispersion contribution has the largest contribution in the $P2_12_12_1$ form but only by ~2 kJ/mol. The repulsion contribution has the lowest contribution in the $P2_1$ form. The $P2_1$ form also has the higher polarization contribution. The $P2_1/c$ form presents again a higher number of generated polymorphs than either of the two forms (33 vs 12 or 4), and it also has the largest range of energies for these forms (~40 kJ/mol range, compared to ~20 and ~7 kJ/mol range for the other two forms). Figure 75 shows the representation of the crystal packing structural forms for the three space groups.

Table 29: Results from the PROM and PIXEL modules (energy in kJ/mol) for 1,6-dibromo-2,5-dimethylcorannulene.

2Br2Me cor	$P2_1$	$P2_1/c$	$P2_12_12_1$
coulomb	-59	-53	-64
dispersion	-237.2	-242	-244.2
repulsion	156.5	167.8	164.1
polarization	-34.4	-40.4	-40

Total (coulomb+dispersion+ repulsion+polarization)	-174.1	-167.6	-184
6-exp (equation 37)	-166.8	-168.3	-171.7
# of polymorphs generated	12	33	4
Range of energy of predicted polymorphs	-152.1,-171.8	-133.6,-172.3	-164.1,-171.7

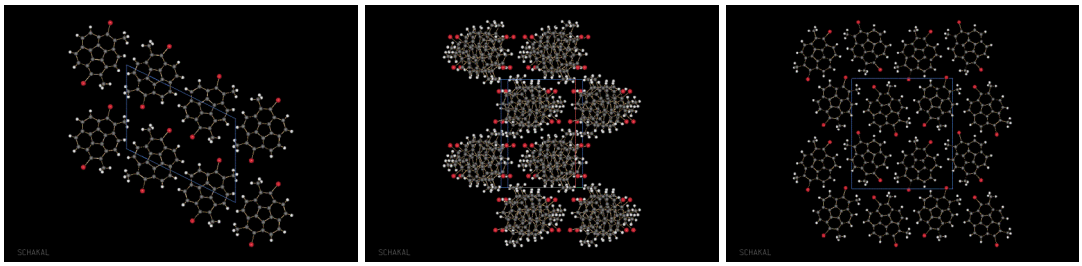


Figure 75: Crystal structures of 1,6-dibromo-2,5-dimethylcorannulene predicted by PROM and used with PIXEL. (a) P21 space group. (b) P21/c space group. (c) P212121 space group.

1,2,5,6-tetrabromocorannulene is observed experimentally to adopt a $Pna2_1$ packing arrangement. Analyzing the theoretical results in terms of the total of the different contributions, one can observe that the $P2_12_12_1$ is the preferred form, by more than ~20 kJ/mol compared to both other forms. Comparing the various individual contributions, one finds that the dispersion has the greatest contribution in the same $P2_12_12_1$ form. $P2_1$ and $P2_1/c$ presents similar values in all energy contributions. $P2_1/c$ has a higher number of generated polymorphs than the other two forms (29 vs. 5 or 3) but it also has the largest range of energies for these forms (~50 kJ/mol vs. ~20 kJ/mol). Figure 76 shows the represented crystal packing structural forms for the three space groups.

Table 30: Results from the PROM and PIXEL modules (energy in kJ/mol) for 1,2,5,6-tetrabromocorannulene.

4Br cor	P21	P21c	P212121
coulomb	-53.8	-57	-64
dispersion	-248.3	-240.5	-255.8
repulsion	170.6	179	171.6
polarization	-40.5	-41.4	-47.7
Total (coulomb+dispersion+	-172	-169.9	-195.9

repulsion+polarization)			
6-exp (equation 37)	-171.1	-171.2	-176.1
# of polymorphs generated	5	29	3
Range of energy of predicted polymorphs	-161.9,-180.3	-133.1,81.9	-165.4,-186.2

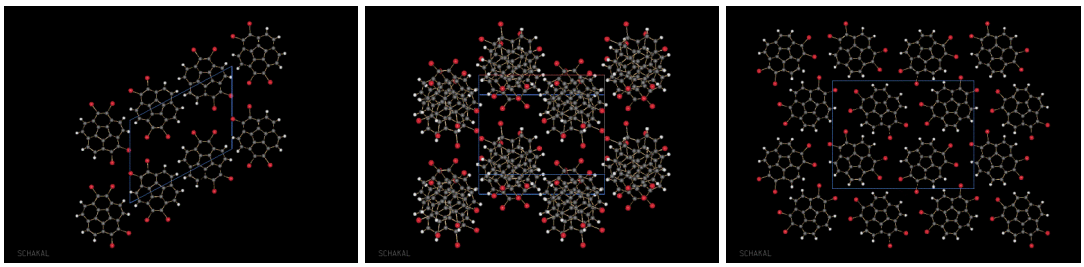


Figure 76: Crystal structures of 1,2,5,6-tetrabromocorannulene predicted by PROM and used with PIXEL. (a) P21 space group. (b) P21/c space group. (c) P212121 space group.

One can further analyze the above polymorphic predictions using the Pixel analysis tool, introduced earlier. The Pixel analysis tool relies on the quantum mechanical charge distribution. The analysis was carried out on the crystal structure conformations at the MP2/cc-pVDZ level of theory. The results of the Pixel analysis using this QM information is provided for the three functionalized corannulenes in Table 31 and represented on Figure 77.

Table 31. Pixel energetic analysis for 1,2,5,6-tetramethylcorannulene, **9**, 1,6-dibromo-2,5-dimethylcorannulene, **11**, and 1,2,5,6-tetrabromocorannulene, **10** (energies in kJ/mol).

	Coulomb E.	Polarization E.	Dispersion E.	Repulsion E.	Total E.
9	-26	-17.2	-208.5	96.6	-155.1
11	-35.3	-19	-255.1	153	-156.5
10	-46.8	-20.4	-287.2	181.9	-172.6

From the data, we observe that the main stabilizing contribution arises from the dispersion energy, becoming increasingly important with the addition of bromine atoms. From the packing arrangement, it is understandable how going from a disordered 1,2,5,6-tetramethylcorannulene, to the columnar stacked 1,6-dibromo-2,5-dimethylcorannulene (neighboring columns in opposite directions) to the columnar structure of 1,2,5,6-tetrabromocorannulene (neighboring columns in the same

direction), the bowl surface available for dispersion interaction becomes larger as well as the corresponding energetic contribution.

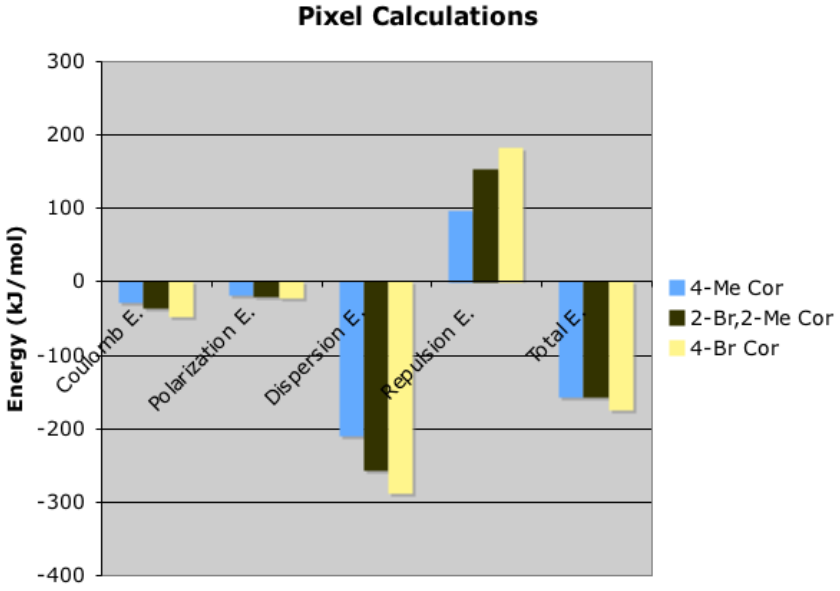


Figure 77: Energy contribution calculated using the Pixel module.

5.5 Conclusion

Quantum chemistry methods were used to investigate corannulene and functionalized corannulene to better understand the behavior of the theoretical methods, as well as supplement the experimental findings. Effects of substitutents in several specific cases were discussed in terms of their electronic structure and property differences.

The application of developed methodology, including continuum solvent method and crystal packing analysis, was applied to polycyclic aromatic hydrocarbon compounds, demonstrating the general application and interpretability for real organic chemistry problems. Solvation studies on the parent corannulene are the beginning of a more general study on the effect of solvent of the family of substituted corannulene. Further investigation can now be carried out to more fully understand the effects of solvation on these interesting structures.

The use of the crystallographic analysis tools combined with the QM electronic structure information was used to explain the tendency of substituted corannulenes to adopt certain space groups. In general, it is worth noting that the polymorph prediction made by PROM module is entirely based on empirical force

fields and on pure geometrical operations. The OPiX package is also empirical, and not designed to explain why a certain molecule chooses a specific packing motif in a crystal, but provides a rough breakdown of energetic contributions and their relative magnitude in the total inter-molecular binding. As such, the set of analyzed results is primarily used as a guide to the experimental information, and not as a quantitative analysis of the molecular structure and packing patterns. In this work, we have analyzed two different classes of systems, substituted benzenes and substituted corannulenes. In the first set, our theoretical predictions of crystal packing were in very good agreement with experimental results. In the second case, the theoretical predictions are much less clear, since the differences in energies between the different polymorph is too small to provide definitive conclusions.

Corannulene is a promising molecule in several domains as biology, semi conductor engineering, metallo-chemistry, as just a few examples. The synthesis of several substituted derivatives of corannulene has recently attracted enormous interest due to their peculiar fluorescence properties.[140] By extending the conjugation of corannulene, the wavelength of absorption and emission can be modified.[146] The aim of the present study was to provide theoretical framework of tools that can be used in conjunction with experimental results to better understand the mechanism of such processes, as well as provide insight into particular substituents that could be used for optimizing certain molecular properties.

6 Conclusion and Perspective

The thesis work involved investigation of QM-based theoretical models for predicting structure and properties of organic molecules in their environment with accuracy. The goals included understanding of complex processes during solvation or crystallization. Such capabilities have impact in life science, material science as well as chemistry and physics in general. The presented work describes the design, implementation, and the application of solvation and crystallization models based on quantum mechanics theory. The first part of the thesis describes solvation processes, used to predict accurate free energy of solvation and properties of molecules in solvent environment from first principles. The second part of the thesis describes methodology for carrying out solid state analysis and application to processes in organic chemistry.

The solvation method presented, COSab, has been successfully used on small organic molecules to predict solvation properties using three different methods to consider possible differences of treatment of the solute. Improvements have been done to the previous implementations of the algorithm, allowing a better accuracy in the description of the cavity and the electrostatic. The inclusion of non-electrostatic effects has been taken into account, which was one of the major drawbacks of the previous model. The model has been shown to describe complexes properties, such as acid dissociation, and the ability to generate σ-profile for fine description of polarizability of molecules and associated electronic features in solvent.

The OPiX package has been introduced and validated on known systems. Application to a series of substituted corannulene molecules sheds light on questions raised on experimental data. The results given by the PIXEL module, in particular, help better understand intermolecular forces in the crystal environment, with a view of the different contribution involved. The complex nature of intermolecular forces involved in crystallization, as predicted by the OPiX model, is not entirely explained in the case of substituted corannulene packing. This is potentially related to the inability to model experimental conditions in this case. However, in general, the use of such tools is shown to improve the understanding of packing structure and properties of molecules.

While there are a variety of computational models for description of molecules in a solvent environment, there is of yet no universal solution. The COSab method presented in this work is part of this community effort to establish models that are able to accurately describe such complex processes. Even more limited are computational models for carrying out crystal analysis, due to the complex nature of the problem and the ability to mimic the experimental conditions. The models presented in this thesis work forge this area by demonstrating one of the more accurate models available, based on QM data.

Further work should involve improvements in the inclusion of non-electrostatic effects in the solvation model. In the current model, the inclusion of such effects is carried out in an iterative manner, as coupling between QM and post-processing statistical mechanics code. As of yet, there is not a direct self-consistent way to include such effects in quantum mechanic Hamiltonian. Moreover, further work is warranted to the improvement in the description of intermolecular forces in the crystalline environment, and the ability to more accurately predict and understand polymorphism.

7 References

(1) Reichardt, C. **2003**, 629.
(2) V. Anslyn, E.; A. Dougherty, D. **2006**, 1095.
(3) Kapustinsky, A.; Weselowsky, B. In *Z Phys Chem B-Chem E* 1933; Vol. 22, p 261-266.
(4) Gavezzotti, A. **2007**, 448.
(5) GAVEZZOTTI, A. In *Struct Chem* 2005; Vol. 16, p 177-185.
(6) Shipley Rowlinson, J. **2002**, 333.
(7) Born, M. In *Z Phys* 1920; Vol. 1, p 45-48.
(8) Kirkwood, J. G. *The Journal of Chemical Physics* **1934**.
(9) Onsager, L. In *Journal of the American Chemical Society* 1936; Vol. 58, p 1486-1493.
(10) Cramer, C. J.; Truhlar, D. G. In *Chem Rev* 1999; Vol. 99, p 2161-2200.
(11) Chapman, D. L. *Phil.Mag.* **1913**, *25*, 475.481.
(12) Gouy, M. *J.Phys.* **1910**, *9*, 457-468.
(13) TOMASI, J.; PERSICO, M. In *Chem Rev* 1994; Vol. 94, p 2027-2094.
(14) Tomasi, J.; Miertus, S.; E, S. *Chem Phys* **1981**, *55*, 117-129.
(15) KLAMT, A.; SCHUURMANN, G. In *J Chem Soc Perk T 2* 1993, p 799-805.
(16) Baldridge, K.; KLAMT, A. In *J Chem Phys* 1997; Vol. 106, p 6622-6633.
(17) Christiansen, O.; Mikkelsen, K. V. In *J Chem Phys* 1999; Vol. 110, p 1365-1375.
(18) Christiansen, O.; Mikkelsen, K. V. In *J Chem Phys* 1999; Vol. 110, p 8348-8360.
(19) Baldridge, K. K.; Jonas, V.; Bain, A. D. In *J Chem Phys* 2000; Vol. 113, p 7519-7529.
(20) Baldridge, K. K.; Jonas, V. In *J Chem Phys* 2000; Vol. 113, p 7511-7518.
(21) Peverati, R.; Baldridge, K. *pubs.acs.org* **2008**.
(22) Tomasi, J.; Mennucci, B.; CAMMI, R. In *Chem Rev* 2005; Vol. 105, p 2999-3093.
(23) FLORIS, F.; TOMASI, J. In *Journal of Computational Chemistry* 1989; Vol. 10, p 616-627.
(24) STILL, W. C.; TEMPCZYK, A.; HAWLEY, R. C.; HENDRICKSON, T. In *Journal of the American Chemical Society* 1990; Vol. 112, p 6127-6129.
(25) CRAMER, C. J.; TRUHLAR, D. G. In *Journal of the American Chemical Society* 1991; Vol. 113, p 8305-8311.
(26) ANDZELM, J.; KOLMEL, C.; KLAMT, A. In *J Chem Phys* 1995; Vol. 103, p 9312-9320.
(27) Ben-Naim, A.; Marcus, Y. J. *J. Chem. Phys.* **1984**, *81*, 2016-2027.
(28) STONE, A. J. In *Chem Phys Lett* 1981; Vol. 83, p 233-239.
(29) GAVEZZOTTI, A. In *Crystengcomm* 2003; Vol. 5, p 429-438.
(30) GAVEZZOTTI, A. In *J Phys Chem B* 2003; Vol. 107, p 2344-2353.
(31) Isaakovich Kitaĭgorodskiĭ, A. **1973**, 553.
(32) MADDOX, J. In *Nature* 1988; Vol. 335, p 201-201.
(33) GAVEZZOTTI, A.; Filippini, G. In *Journal of the American Chemical Society* 1996; Vol. 118, p 7153-7157.
(34) GAVEZZOTTI, A. In *Accounts Chem Res* 1994; Vol. 27, p 309-314.

(35) GAVEZZOTTI, A. In *Journal of the American Chemical Society* 1991; Vol. 113, p 4622-4629.
(36) GAVEZZOTTI, A.; Filippini, G. In *Synthetic Met* 1991; Vol. 40, p 257-266.
(37) Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. In *Chem Phys Lett* 2008; Vol. 456, p 105-109.
(38) GAVEZZOTTI, A. In *J Phys Chem B* 2002; Vol. 106, p 4145-4154.
(39) KLAMT, A. *The Journal of Physical Chemistry* **1995**.
(40) KLAMT, A.; Jonas, V. *The Journal of Chemical Physics* **1996**.
(41) BONDI, A. In *J Phys Chem-Us* 1964; Vol. 68, p 441-&.
(42) KLAMT, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**.
(43) Gregerson, L. N.; BALDRIDGE, K. K. In *Helv Chim Acta* 2003; Vol. 86, p 4112-4132.
(44) STONE, A. J.; ALDERTON, M. In *Mol Phys* 1985; Vol. 56, p 1047-1064.
(45) SCHMIDT, M. W.; BALDRIDGE, K. K.; BOATZ, J. A.; ELBERT, S. T.; GORDON, M. S.; JENSEN, J. H.; KOSEKI, S.; MATSUNAGA, N.; NGUYEN, K. A.; SU, S. J.; WINDUS, T. L.; DUPUIS, M.; MONTGOMERY, J. A. In *Journal of Computational Chemistry* 1993; Vol. 14, p 1347-1363.
(46) POPLE, J. A.; NESBET, R. K. In *J Chem Phys* 1954; Vol. 22, p 571-572.
(47) ROOTHAAN, C. C. J. In *Rev Mod Phys* 1951; Vol. 23, p 69-89.
(48) POPLE, J. A.; BINKLEY, J. S.; SEEGER, R. In *Int J Quantum Chem* 1976, p 1-19.
(49) Moller, C.; Plesset, M. S. In *Physical Review* 1934; Vol. 46, p 0618-0622.
(50) BECKE, A. D. In *J Chem Phys* 1993; Vol. 98, p 1372-1377.
(51) BECKE, A. D. In *J Chem Phys* 1993; Vol. 98, p 5648-5652.
(52) PARR, R. G. In *Int J Quantum Chem* 1990; Vol. 37, p 327-347.
(53) LEE, C. T.; YANG, W. T.; PARR, R. G. In *Phys Rev B* 1988; Vol. 37, p 785-789.
(54) BECKE, A. D. In *Int J Quantum Chem* 1983; Vol. 23, p 1915-1922.
(55) KOHN, W.; SHAM, L. J. In *Physical Review* 1965; Vol. 140, p 1133-&.
(56) Pascualahuir, J. L.; SILLA, E. *Journal of Computational Chemistry* **1990**, *11*, 1047-1060.
(57) Klamt, A. **2005**, 234.
(58) B. Lipkowitz, K.; B. Boyd, D.
(59) BECKE, A. D. *Physical Review A* **1988**, *38*, 3098-3100.
(60) Kemp, D. D.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 4885–4894.
(61) Eckert, F.; KLAMT, A. In *AIChE Journal* 2002; Vol. 48, p 369-385.
(62) KLAMT, A.; Eckert, F.; Hornig, M. *Journal of Computer-Aided Molecular Design* **2001**.
(63) KLAMT, A.; Eckert, F. *Fluid Phase Equilibria* **2000**.
(64) WINGET, P.; LIOTARD, D. A.; CRAMER, C. J.; TRUHLAR, D. G. *Journal of Computational Chemistry* **2000**.
(65) Cramer, C. J.; Truhlar, D. G. In *Accounts Chem Res* 2008; Vol. 41, p 760-768.

(66) Perdew, J. P. *Physical Review B* **1986**, *33*, 8822-8824.
(67) Weast…, R. C. *osti.gov* **1975**.
(68) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. In *J Phys Chem B* 2008; Vol. 112, p 8651-8655.
(69) Pliego, J. R. In *Chem Phys Lett* 2003; Vol. 367, p 145-149.
(70) Chipman, D. M. In *Journal of Physical Chemistry A* 2002; Vol. 106, p 7413-7422.
(71) Kelly, C. P.; CRAMER, C. J.; TRUHLAR, D. G. *Journal of Physical Chemistry A* **2006**.
(72) KLAMT, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. In *Journal of Physical Chemistry A* 2003; Vol. 107, p 9380-9386.
(73) Kaminski, G. A. In *J Phys Chem B* 2005; Vol. 109, p 5884-5890.
(74) Eckert, F.; KLAMT, A. *J Comput Chem* **2006**.
(75) POTGIETER, J. H. In *J Chem Educ* 1991; Vol. 68, p 280-281.
(76) Dinadayalane, T. C.; Priyakumar, U. D.; Sastry, G. N. In *Journal of Physical Chemistry A* 2004; Vol. 108, p 11433-11448.
(77) VANTAMELEN, E. E.; PAPPAS, S. P. In *Journal of the American Chemical Society* 1963; Vol. 85, p 3297-&.
(78) WILZBACH, K. E.; RITSCHER, J. S.; KAPLAN, L. In *Journal of the American Chemical Society* 1967; Vol. 89, p 1031-&.
(79) KATZ, T. J.; WANG, E. J.; ACTON, N. In *Journal of the American Chemical Society* 1971; Vol. 93, p 3782-&.
(80) KATZ, T. J.; ACTON, N. In *Journal of the American Chemical Society* 1973; Vol. 95, p 2738-2739.
(81) BILLUPS, W. E.; HALEY, M. M. In *Angew Chem Int Edit* 1989; Vol. 28, p 1711-1712.
(82) Koopmans, T. In *Physica* 1934; Vol. 1, p 104-113.
(83) Metrangolo, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. In *Angew Chem Int Edit* 2008; Vol. 47, p 6114-6127.
(84) GAVEZZOTTI, A. In *Mol Phys* 2008; Vol. 106, p 1473-1485.
(85) Thalladi, V. R.; Weiss, H. C.; Blaser, D.; Boese, R.; Nangia, A.; Desiraju, G. R. In *Journal of the American Chemical Society* 1998; Vol. 120, p 8702-8710.
(86) Pauling, L. **1960**.
(87) Lennard-Jones, J. E. In *P Phys Soc* 1931; Vol. 43, p 461-482.
(88) London, F. *Transactions of the Faraday Society* **1937**.
(89) Dunitz, J. D.; Gavezzotti, A. In *Angew Chem Int Edit* 2005; Vol. 44, p 1766-1787.
(90) GAVEZZOTTI, A. In *Crystengcomm* 2003; Vol. 5, p 439-446.
(91) Destri, S.; Pasini, M.; Porzio, W.; GAVEZZOTTI, A.; Filippini, G. In *Cryst Growth Des* 2003; Vol. 3, p 257-262.
(92) Dunitz, J. D.; GAVEZZOTTI, A. In *Helv Chim Acta* 2002; Vol. 85, p 3949-3964.
(93) Dunitz, J. D.; Filippini, G.; GAVEZZOTTI, A. In *Helv Chim Acta* 2000; Vol. 83, p 2317-2335.
(94) GAVEZZOTTI, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P. In *Chem-Eur J* 1997; Vol. 3, p 893-899.
(95) GAVEZZOTTI, A. In *Synlett* 2002, p 201-214.
(96) GAVEZZOTTI, A. In *Model Simul Mater Sc* 2002; Vol. 10, p R1-R29.
(97) Gavezzotti, A. **1997**, 248.

(98) GAVEZZOTTI, A. In *Acta Crystallogr B* 1996; Vol. 52, p 201-208.
(99) GAVEZZOTTI, A.; Filippini, G. In *Journal of the American Chemical Society* 1995; Vol. 117, p 12299-12305.
(100) FILIPPINI, G.; GAVEZZOTTI, A. In *Acta Crystallogr B* 1993; Vol. 49, p 868-880.
(101) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C, Pople, J. A. *Gaussian, Inc., Wallingford CT* **2004**.
(102) Robertson, J. M. In *Proc R Soc Lon Ser-A* 1935; Vol. 150, p 0106-0128.
(103) TROTTER, J. In *Acta Crystallogr* 1960; Vol. 13, p 86-95.
(104) Bolte M., M. G. *Private Communication* **2002**.
(105) SHIMONI, L.; GLUSKER, J. P. In *Struct Chem* 1994; Vol. 5, p 383-397.
(106) DESIRAJU, G. R. In *Chem Commun* 1997, p 1475-1482.
(107) Howard, J. A. K.; Hoy, V. J.; OHagan, D.; Smith, G. T. In *Tetrahedron* 1996; Vol. 52, p 12613-12622.
(108) Boese, R.; Kirchner, M. T.; Dunitz, J. D.; Filippini, G.; GAVEZZOTTI, A. In *Helv Chim Acta* 2001; Vol. 84, p 1561-1577.
(109) STONE, A. J.; Price, S. L. In *J Phys Chem-Us* 1988; Vol. 92, p 3325-3335.
(110) PERTSIN, A. I.; IVANOV, Y. P.; KITAIGORODSKI, A. I. In *Kristallografiya+* 1981; Vol. 26, p 115-124.
(111) J. Pertsin, A.; London, S.; Limited; Isaakovich Kitaigorodski, A. **1987**, 397.
(112) BONDI, A. In *Chem Rev* 1967; Vol. 67, p 565-&.
(113) BONDI, A. *J. Chem. Eng. Data* **1963**, *8*, 371-381.
(114) CHICKOS, J. S.; ANNUNZIATA, R.; LADON, L. H.; HYMAN, A. S.; LIEBMAN, J. F. In *J Org Chem* 1986; Vol. 51, p 4311-4314.
(115) Zhao, Y.; Truhlar, D. G. In *Theor Chem Acc* 2008; Vol. 120, p 215-241.
(116) DUNNING, T. H. In *J Chem Phys* 1989; Vol. 90, p 1007-1023.
(117) GAVEZZOTTI, A. In *J Chem Theory Comput* 2005; Vol. 1, p 834-840.
(118) Bernstein, M. P.; Sandford, S. A.; Allamandola, L. J.; Gillette, J. S.; Clemett, S. J.; Zare, R. N. In *Science* 1999; Vol. 283, p 1135-1138.
(119) Grimmer, G. **1983**.
(120) G. Harvey, R. **1997**, 682.

(121) KROTO, H. W.; HEATH, J. R.; OBRIEN, S. C.; CURL, R. F.; SMALLEY, R. E. In *Nature* 1985; Vol. 318, p 162-163.
(122) BARTH, W. E.; LAWTON, R. G. In *Journal of the American Chemical Society* 1971; Vol. 93, p 1730-&.
(123) HANSON, J. C.; NORDMAN, C. E. In *Acta Crystallogr B* 1976; Vol. 32, p 1147-1153.
(124) Hedberg, L.; Hedberg, K.; Cheng, P. C.; Scott, L. T. In *Journal of Physical Chemistry A* 2000; Vol. 104, p 7689-7694.
(125) Seiders, T. J.; Baldridge, K. K.; Grube, G. H.; Siegel, J. S. In *Journal of the American Chemical Society* 2001; Vol. 123, p 517-525.
(126) Scott, L. T.; Bronstein, H. E.; Preda, D. V.; Ansems, R. B. M.; Bratcher, M. S.; Hagen, S. In *Pure Appl Chem* 1999; Vol. 71, p 209-219.
(127) Scott, L. T.; HASHEMI, M. M.; Bratcher, M. S. In *Journal of the American Chemical Society* 1992; Vol. 114, p 1920-1921.
(128) Baldridge, K. K.; Siegel, J. S. In *Theor Chem Acc* 1997; Vol. 97, p 67-71.
(129) HADDON, R. C.; Scott, L. T. In *Pure Appl Chem* 1986; Vol. 58, p 137-142.
(130) HADDON, R. C. In *Science* 1993; Vol. 261, p 1545-1550.
(131) HADDON, R. C. In *Journal of the American Chemical Society* 1990; Vol. 112, p 3385-3389.
(132) HADDON, R. C. In *Accounts Chem Res* 1988; Vol. 21, p 243-249.
(133) RABIDEAU, P. W.; MARCINOW, Z.; SYGULA, R.; SYGULA, A. In *Tetrahedron Lett* 1993; Vol. 34, p 6351-6354.
(134) Petrukhina, M. A.; Andreini, K. W.; Mack, J.; Scott, L. T. In *J Org Chem* 2005; Vol. 70, p 5713-5716.
(135) Grimme, S.; Antony, J.; Schwabe, T.; Mueck-Lichtenfeld, C. In *Org Biomol Chem* 2007; Vol. 5, p 741-758.
(136) Seiders, T. J.; Baldridge, K. K.; Elliott, E. L.; Grube, G. H.; Siegel, J. S. In *Journal of the American Chemical Society* 1999; Vol. 121, p 7439-7440.
(137) BURGI, H. B.; Dunitz, J. D. In *Journal of the American Chemical Society* 1987; Vol. 109, p 2924-2926.
(138) BURGI, H. B.; Dunitz, J. D.; SHEFTER, E. In *Journal of the American Chemical Society* 1973; Vol. 95, p 5065-5067.
(139) Hayama, T.; Baldridge, K. K.; Wu, Y.; Linden, A.; Siegel, J. S. In *Journal of the American Chemical Society* 2008; Vol. 130, p 1583-1591.
(140) Wu, Y.; Bandera, D.; Maag, R.; Linden, A.; Baldridge, K. K.; Siegel, J. S. In *Journal of the American Chemical Society* 2008; Vol. 130, p 10729-10739.
(141) Hayama, T. **2005**.
(142) Sevryugina, Y.; Rogachev, A. Y.; Jackson, E. A.; Scott, L. T.; Petrukhina, M. A. In *J Org Chem* 2006; Vol. 71, p 6615-6618.
(143) Hayama, T.; Wu, Y.; Linden, A.; Baldridge, K. K.; Siegel, J. S. In *Journal of the American Chemical Society* 2007; Vol. 129, p 12612-+.
(144) Siegel, J. S. *Private Communition* **2005**.
(145) Wu, Y.; Siegel, J. S. In *Chem Rev* 2006; Vol. 106, p 4843-4867.
(146) Mack, J.; Vogel, P.; Jones, D.; Kaval, N.; Sutton, A. In *Org Biomol Chem* 2007; Vol. 5, p 2448-2452.

8 Appendix

Molecules	ΔG hydr exp kcal/mol	DBLCAV				DMULTI				DCOSMO					surface area of cavity (A**2)	volume of cavity (A**3)
		ΔE_{elect} (kcal/mol)	OCE (kcal/mol)	Dielectric Energy (a.u.)	Sum of polarization charges	ΔE_{elect} (kcal/mol)	OCE (kcal/mol)	Dielectric Energy (a.u.)	Sum of polarization charges	ΔG (kcal/mol)	OCE (kcal/mol)	OCE DCOSMO-RS (kcal/mol)	Dielectric Energy (a.u.)	Sum of polarization charges		
methane	1.97	-0.53	0.0377	-0.00075	0.01990	-0.40	-0.0839	-0.00060	0.00066	-0.54	0.0384	0.0384	-0.00076	0.01990	199.93	256.01
ethane	1.81	-0.86	0.0510	-0.00077	0.02482	-0.68	-0.1801	-0.00061	0.00059	0.38	0.2496	0.0932	-0.00107	0.02491	276.54	396.54
propane	1.97	-0.49	0.0613	-0.00093	0.02849	-0.45	-0.2860	-0.00079	0.00059	0.75	0.3310	0.1231	-0.00133	0.02860	348.46	544.47
butane	2.12	-0.56	0.0712	-0.00107	0.03153	-0.50	-0.3449	-0.00092	0.00060	0.93	0.4154	0.1539	-0.00158	0.03165	419.5	692.77
isobutane	2.31	-0.61	0.0686	-0.00116	0.03105	-0.56	-0.4018	-0.00107	0.00072	0.86	0.3760	0.1410	-0.00161	0.03117	414.17	699.81
pentane	2.33	-0.50	0.0783	-0.00121	0.03436	-0.45	-0.4232	-0.00106	0.00062	1.17	0.4956	0.1834	-0.00184	0.03449	490.15	838.91
2-methyl-butane	2.38	-0.67	0.0761	-0.00126	0.03337	-0.59	-0.4374	-0.00115	0.00067	1.01	0.4478	0.1665	-0.00181	0.03349	474.82	839.29
2,2-dimethyl-propane	2.58	-0.82	0.0694	-0.00142	0.03291	-0.76	-0.5095	-0.00140	0.00093	0.87	0.3820	0.1454	-0.00186	0.03303	467.98	851.18
hexane	2.53	-0.50	0.0852	-0.00135	0.03705	-0.47	-0.5045	-0.00120	0.00064	1.40	0.5761	0.2130	-0.00209	0.03720	561.15	986.19
octane	2.89	-0.88	0.0976	-0.00163	0.04208	-0.85	-0.6701	-0.00148	0.00075	-0.90	0.0994	0.0994	-0.00166	0.04207	703.42	1281.7
cyclopropane	0.75	-2.05	0.0440	-0.00253	0.02346	-1.87	-0.1449	-0.00237	0.00085	-0.65	-0.1992	-0.0939	-0.00202	0.02358	316.82	481.04
cyclopentane	1.2	-2.40	0.0766	-0.00105	0.03116	-2.15	-0.3697	-0.00094	0.00038	0.19	0.4450	0.1650	-0.00160	0.03131	427.82	736.73
methyl-cyclopentane	1.6	-0.81	0.0853	-0.00122	0.03364	-0.71	-0.4409	-0.00110	0.00049	1.05	0.5106	0.1889	-0.00185	0.03378	494	891.42
cyclohexane	1.23	-2.90	0.0813	-0.00100	0.03441	-2.55	-0.4354	-0.00088	0.00034	1.07	0.5334	0.1943	-0.00170	0.03456	469.09	852.05
cyclooctane	0.85	-0.75	0.0810	-0.00121	0.03797	-0.63	-0.5681	-0.00115	0.00034	-0.77	0.0824	0.0824	-0.00124	0.03796	579.37	1144.68
ethene	1.27	-1.82	-0.0257	-0.00270	0.02130	-1.78	-0.2058	-0.00276	0.00053	-0.89	-0.3432	-0.1842	-0.00203	0.02135	258.32	363.58
propene	1.27	-1.84	-0.0028	-0.00298	0.02550	-1.79	-0.2992	-0.00304	0.00050	-0.64	-0.2980	-0.1491	-0.00226	0.02562	332.3	507.71
butene	1.38	-1.88	0.0165	-0.00299	0.02882	-1.79	-0.3719	-0.00303	0.00053	-0.39	-0.2003	-0.0937	-0.00239	0.02895	403.63	656.05
isobutene	1.22	-2.38	0.0199	-0.00326	0.02824	-2.28	-0.3810	-0.00331	0.00062	-0.73	-0.2479	-0.1193	-0.00249	0.02844	398.76	657.44
cis-but-2-ene		-2.74	0.0083	-0.00299	0.02882	-2.57	-0.3602	-0.00301	0.00053	-0.93	-0.2277	-0.1066	-0.00235	0.02902	398.68	645.63
pentene	1.66	-1.78	0.0305	-0.00312	0.03186	-1.66	-0.4400	-0.00316	0.00054	-0.13	-0.1307	-0.0592	-0.00261	0.03198	473.69	802.7
2-methyl-but-2-ene	1.31	-2.53	0.0492	-0.00300	0.03125	-2.38	-0.4012	-0.00294	0.00059	0.05	-0.0950	-0.0426	-0.00247	0.03155	464.27	793.68
hexene	1.68	-1.65	0.0412	-0.00329	0.03469	-1.52	-0.5196	-0.00332	0.00055	0.07	-0.0622	-0.0276	-0.00284	0.03482	545.14	951.31
octene	2.17	-1.45	0.0567	-0.00356	0.03989	-1.27	-0.6861	-0.00360	0.00061	0.45	0.0865	0.0362	-0.00331	0.04003	687.71	1247.09
cyclopentene	0.56	-1.65	0.0258	-0.00262	0.02932	-1.63	-0.3539	-0.00262	0.00038	-0.11	-0.0967	-0.0434	-0.00218	0.02952	413.24	696.12
cyclohexene	0.37	-1.61	0.0236	-0.00281	0.03224	-1.67	-0.4857	-0.00288	0.00023	0.04	-0.0716	-0.0320	-0.00239	0.03244	462.67	820.83
methyl-cyclohexene	0.67	-1.59	0.0582	-0.00274	0.03440	-1.61	-0.4997	-0.00271	0.00038	0.35	0.0851	0.0348	-0.00251	0.03471	531.41	973.79
propadiene		-4.82	-0.0792	-0.00417	0.02245	-2.63	-0.2350	-0.00429	0.00052	-1.64	-0.5830	-0.3407	-0.00299	0.02259	316.99	470.4

buta-1,3-diene	0.59	-2.63	-0.0396	-0.00443	0.02626	-2.62	-0.3810	-0.00462	0.00052	-1.39	-0.5624	-0.3092	-0.00327	0.02639	387.43	618.83
isopropene	0.68	-2.43	-0.0236	-0.00442	0.02842	-2.51	-0.4360	-0.00459	0.00051	-0.91	-0.4950	-0.2556	-0.00330	0.02862	445.51	762.09
penta-1,4-diene	0.94	-2.76	-0.0202	-0.00447	0.02933	-2.71	-0.4483	-0.00465	0.00046	-1.23	-0.5079	-0.2640	-0.00340	0.02946	456.08	765.46
ethyne	-0.01	-3.95	-0.1830	-0.00642	0.01695	-4.06	-0.3894	-0.00700	0.00061	-3.69	-0.5399	-0.4707	-0.00582	0.01693	243.09	339.36
propyne	-0.31	-3.87	-0.1265	-0.00668	0.02181	-3.87	-0.3317	-0.00703	0.00042	-3.16	-0.9217	-0.6466	-0.00500	0.02183	318.89	476.19
but-1-yne	-0.16	-3.83	-0.0857	-0.00645	0.02568	-3.89	-0.3935	-0.00675	0.00046	-2.67	-0.9391	-0.5835	-0.00455	0.02576	388.78	618.79
but-2-yne		-3.77	-0.0630	-0.00620	0.02600	-3.72	-0.2616	-0.00630	0.00048	-2.45	-0.8125	-0.5001	-0.00439	0.02617	393.93	608.06
pent-1-yne	0.01	-3.98	-0.0556	-0.00651	0.02906	-4.05	-0.4705	-0.00683	0.00046	-2.45	-0.9331	-0.5382	-0.00454	0.02912	460.99	768.92
hex-1-yne	0.29	-4.11	-0.0392	-0.00665	0.03196	-4.18	-0.5453	-0.00696	0.00044	-2.26	-0.9173	-0.5013	-0.00462	0.03202	531.15	916.64
oct-1-yne	0.71	-4.28	-0.0215	-0.00692	0.03745	-4.36	-0.7183	-0.00725	0.00058	-1.83	-0.8592	-0.4324	-0.00488	0.03752	674.12	1214.59
but-1-en-3-yne	0.04	-4.08	-0.1322	-0.00662	0.02294	-4.16	-0.3730	-0.00703	0.00042	-3.02	-0.9654	-0.6375	-0.00480	0.02303	373.8	583.57
methanol	-5.09	-6.24	-0.0936	-0.01043	0.01818	-6.18	-0.2780	-0.01070	0.00008	-6.31	-1.6478	-1.2330	-0.00740	0.01825	242.35	327.96
ethanol	-4.96	-6.40	-0.0893	-0.01020	0.02333	-6.30	-0.3959	-0.01049	0.00013	-6.01	-1.9903	-1.2895	-0.00627	0.02349	315.82	471.62
propan-1-ol	-4.84	-6.52	-0.0750	-0.01012	0.02690	-6.39	-0.4731	-0.01042	0.00021	-5.70	-2.2253	-1.2679	-0.00561	0.02707	387.02	619.85
propan-2-ol	-4.76	-6.53	-0.0396	-0.01002	0.02664	-6.41	-0.4850	-0.01030	0.00030	-5.82	-1.8900	-1.1848	-0.00610	0.02682	381.84	621.88
butan-1-ol	-4.72	-6.76	-0.0637	-0.01029	0.02996	-6.59	-0.5475	-0.01059	0.00025	-5.57	-2.3643	-1.2322	-0.00537	0.03017	457.69	767.16
butan-2-ol	-4.59	-6.55	-0.0144	-0.00934	0.02954	-6.38	-0.5359	-0.00958	0.00040	-5.11	-2.0659	-1.0566	-0.00487	0.02980	450.42	769.25
2-methyl-propan-1-ol	-4.52	-6.40	-0.0503	-0.00936	0.02935	-6.24	-0.5703	-0.00967	0.00049	-4.90	-2.3163	-1.0801	-0.00450	0.02958	446.72	769.75
2-methyl-propan-2-ol	-4.51	-6.73	-0.0052	-0.00984	0.02870	-6.55	-0.5315	-0.01009	0.00054	-5.72	-1.8061	-1.1010	-0.00599	0.02889	440.33	775.66
pentan-1-ol	-4.48	-6.98	-0.0549	-0.01040	0.03285	-6.77	-0.6299	-0.01070	0.00028	-5.38	-2.4529	-1.1854	-0.00523	0.03307	527.83	913.04
heptan-1-ol	-4.24	-7.45	-0.0398	-0.01068	0.03813	-7.17	-0.7983	-0.01098	0.00036	-5.07	-2.5615	-1.0864	-0.00514	0.03839	671	1210.64
octan-1-ol	-4.1	-7.68	-0.0323	-0.01084	0.04062	-7.37	-0.8818	-0.01114	0.00041	-4.91	-2.6006	-1.0391	-0.00515	0.04091	742.62	1359.09
prop-2-en-1-ol	-5.03	-6.10	-0.0780	-0.01024	0.02386	-6.16	-0.3799	-0.01051	-0.00002	-5.58	-2.1319	-1.3034	-0.00617	0.02399	368.41	581.78
prop-2-yn-1-ol		-8.46	-0.2208	-0.01283	0.02080	-8.55	-0.4921	-0.01350	-0.00011	-8.11	-2.5759	-1.7503	-0.00831	0.02087	355.9	548.2
cyclohexanol	-5.47	-5.30	-0.0091	-0.00986	0.03260	-5.41	-0.5998	-0.01012	0.00017	-4.39	-2.0239	-1.0955	-0.00548	0.03278	508.5	940.54
methoxy-methane	-1.89	-4.67	0.0693	-0.00657	0.02339	-4.47	-0.1816	-0.00644	0.00021	-4.13	-0.0227	-0.0215	-0.00635	0.02346	319.63	481.8
ethoxy-ethane	-1.69	-5.97	0.0802	-0.00638	0.03176	-5.73	-0.4999	-0.00648	0.00030	-4.55	-0.2383	-0.1837	-0.00543	0.03197	463.82	763.17
1-propoxy-propane	-1.15	-6.46	0.0927	-0.00634	0.03741	-6.17	-0.6862	-0.00652	0.00041	-4.21	-0.2768	-0.1875	-0.00517	0.03765	607.8	1062.22
methoxy-propane	-1.66	-5.50	0.0814	-0.00641	0.03095	-5.21	-0.3803	-0.00634	0.00035	-4.16	-0.2509	-0.1932	-0.00548	0.03112	462.42	770.24
2-methoxy-2-methyl-propane	-2.21	-6.49	0.0574	-0.00691	0.03091	-6.14	-0.5042	-0.00700	0.00068	-2.31	-0.1746	-0.1414	-0.00615	0.03106	501.02	921.13
dimethoxy-methane	-2.93	-7.16	0.0496	-0.00899	0.02542	-6.83	-0.2560	-0.00896	0.00015	-6.35	-0.3286	-0.2847	-0.00827	0.02553	427.57	715.05

1,2-dimethoxy-ethane	-4.83	-5.92	0.0938	-0.01172	0.02826	-6.01	-0.3317	-0.01166	0.00009	-5.25	-0.1317	-0.1206	-0.01124	0.02836	486	858.43
oxirane		-5.33	0.0501	-0.00838	0.01765	-5.13	-0.0326	-0.00820	-0.00002	-4.98	-0.3483	-0.3081	-0.00752	0.01773	285.85	418.4
2-methyl-oxirane		-5.84	0.0861	-0.00864	0.02243	-5.57	-0.1000	-0.00842	0.00009	-5.15	-0.4162	-0.3500	-0.00743	0.02255	356.32	563.33
2-ethyl-oxirane	-2.91	-4.03	0.0996	-0.00831	0.02595	-4.02	-0.1819	-0.00810	0.00011	-3.17	-0.5913	-0.4497	-0.00654	0.02613	426.46	713.42
tetrahydrofurane	-3.47	-6.17	0.0539	-0.00781	0.02711	-5.93	-0.3170	-0.00782	0.00007	-5.54	0.1444	0.1337	-0.00798	0.02716	401.3	671.88
tetrahydropyrane	-3.12	-6.14	0.0859	-0.00694	0.03037	-5.75	-0.3571	-0.00684	0.00003	-5.04	-0.0828	-0.0714	-0.00642	0.03049	451.56	801
dioxolane	-4.1	-7.99	0.0161	-0.01098	0.02290	-7.74	-0.2680	-0.01107	-0.00023	-7.57	-0.2725	-0.2519	-0.01049	0.02296	375.08	610.17
dioxane	-5.05	-8.98	0.0274	-0.01122	0.02673	-8.59	-0.3528	-0.01131	-0.00024	-8.53	0.0141	0.0131	-0.01133	0.02677	427.34	739.14
formaldehyde		-5.05	0.0229	-0.00882	0.01440	-4.94	-0.0672	-0.00880	-0.00024	-5.16	0.0212	0.0212	-0.00909	0.01438	220.11	292.22
ethanal	-3.5	-6.38	0.0786	-0.01100	0.02000	-6.18	-0.0946	-0.01084	-0.00030	-6.17	-0.2624	-0.2435	-0.01032	0.02005	295.65	433.27
propanal	-3.44	-5.01	0.1216	-0.01051	0.02367	-5.07	-0.1624	-0.01034	-0.00018	-4.51	-0.5558	-0.4683	-0.00878	0.02379	363.11	577.23
butanal	-3.18	-5.03	0.1456	-0.01049	0.02719	-5.03	-0.2192	-0.01030	-0.00008	-4.25	-0.6995	-0.5519	-0.00823	0.02733	433.44	725.44
2-methyl-propanal	-2.9	-4.76	0.1425	-0.00997	0.02648	-4.86	-0.2398	-0.00982	0.00004	-3.95	-0.7172	-0.5539	-0.00769	0.02662	427.36	728.56
hexanal	-2.81	-5.17	0.1714	-0.01077	0.03311	-5.25	-0.3829	-0.01058	-0.00000	-3.82	-0.8951	-0.6325	-0.00773	0.03330	575.19	1020.29
octanal	-2.29	-5.32	0.1862	-0.01106	0.03832	-5.42	-0.5505	-0.01087	0.00009	-3.42	-0.9760	-0.6328	-0.00757	0.03857	717.76	1317.17
prop-2-enal	-3.1	-4.86	0.0379	-0.01175	0.02058	-4.86	-0.1576	-0.01173	-0.00028	-4.62	-0.6300	-0.5531	-0.01018	0.02066	349.74	543.66
propanone	-3.85	-7.70	0.0660	-0.01225	0.02308	-7.45	-0.2012	-0.01217	-0.00006	-5.53	-0.2236	-0.2076	-0.01152	0.02314	367.66	584.8
butan-2-one	-3.67	-5.58	0.1046	-0.01163	0.02646	-5.64	-0.2583	-0.01152	0.00007	-4.94	-0.4792	-0.4124	-0.01010	0.02655	432.82	725.55
pentan-2-one	-3.53	-5.59	0.1214	-0.01160	0.02970	-5.66	-0.3386	-0.01151	0.00012	-4.67	-0.6385	-0.5185	-0.00952	0.02980	503.96	872.35
pentan-3-one	-3.41	-5.47	0.1376	-0.01111	0.02968	-5.56	-0.3239	-0.01100	0.00025	-4.46	-0.6186	-0.4952	-0.00896	0.02980	498.38	867.05
3-methyl-butan-2-one	-3.24	-5.43	0.1064	-0.01124	0.02863	-5.50	-0.3206	-0.01113	0.00020	-4.54	-0.6335	-0.5142	-0.00922	0.02874	490.65	874.42
heptan-2-one	-3.04	-5.74	0.1330	-0.01192	0.03533	-5.82	-0.4901	-0.01181	0.00025	-5.89	0.1338	0.1338	-0.01229	0.03532	646.3	1167.25
octan-2-one	-2.88	-5.81	0.1412	-0.01208	0.03792	-5.90	-0.5743	-0.01196	0.00027	-4.11	-0.8847	-0.6345	-0.00901	0.03808	716.92	1314.08
cyclohexanone	-4.6	-5.83	0.1259	-0.01199	0.02978	-5.91	-0.3367	-0.01192	-0.00030	-5.03	-0.4632	-0.3950	-0.01045	0.02985	493.75	898
buta-2,3-dione		-9.08	0.0388	-0.01238	0.02247	-8.82	-0.2555	-0.01243	-0.00037	-8.42	-1.0555	-0.8679	-0.01022	0.02255	445.97	759.32
chinone		-8.31	-0.0447	-0.01573	0.02247	-8.43	-0.3304	-0.01601	-0.00098	-8.19	-1.1995	-1.0381	-0.01342	0.02252	489.13	861.14
acetic acid	-6.7	-8.31	-0.1424	-0.01359	0.01876	-8.22	-0.3261	-0.01391	-0.00042	-8.40	-2.3701	-1.7344	-0.00929	0.01887	330.99	503.74
propanoic acid	-6.47	-8.06	-0.0891	-0.01315	0.02265	-7.97	-0.3782	-0.01344	-0.00028	-7.66	-2.7134	-1.7286	-0.00789	0.02281	397.64	648.12
butanoic acid	-6.35	-8.02	-0.0624	-0.01313	0.02614	-7.92	-0.4525	-0.01342	-0.00021	-7.24	-2.9621	-1.6997	-0.00727	0.02629	468.78	797.01
formic acid methyl ester	-2.78	-6.58	-0.0250	-0.01043	0.01853	-6.48	-0.2051	-0.01057	-0.00019	-6.41	-0.6792	-0.5998	-0.00924	0.01859	332.32	511.91
formic acid ethyl ester	-2.65	-7.16	0.0389	-0.01057	0.02322	-7.07	-0.2835	-0.01068	-0.00016	-6.51	-0.8780	-0.7147	-0.00871	0.02331	405.35	654.27

acetic acid methyl ester	-3.31	-7.48	0.0199	-0.01099	0.02230	-7.38	-0.2627	-0.01107	-0.00011	-6.92	-0.6692	-0.5743	-0.00963	0.02238	404.96	654.85
acetic acid ethyl ester	-3.09	-8.16	0.0713	-0.01119	0.02656	-8.11	-0.3949	-0.01133	-0.00010	-7.19	-0.7992	-0.6484	-0.00937	0.02668	476.59	793.74
acetic acid propyl ester	-2.86	-8.44	0.0966	-0.01119	0.02982	-8.38	-0.4707	-0.01132	-0.00005	-7.07	-0.9176	-0.7046	-0.00901	0.02995	548.01	941.54
propanoic acid ethyl ester	-2.8	-8.04	0.1081	-0.01080	0.02992	-7.99	-0.4723	-0.01093	0.00003	-6.64	-0.9048	-0.6838	-0.00856	0.03007	543.62	937.09
carbonic acid dimethyl ester		-8.69	0.0026	-0.01241	0.02192	-8.63	-0.2953	-0.01258	-0.00012	-8.15	-0.9823	-0.8245	-0.01058	0.02202	444.11	733.27
carbonic acid diethyl ester		-9.72	0.0984	-0.01278	0.03027	-9.69	-0.4714	-0.01292	-0.00008	-8.32	-1.2191	-0.9242	-0.01011	0.03044	586.65	1009.54
benzene	-0.87	-4.52	-0.0454	-0.00551	0.02741	-3.31	-0.3737	-0.00565	0.00033	-2.65	-0.6961	-0.3980	-0.00393	0.02761	433.67	743.16
toluene	-0.82	-3.43	-0.0223	-0.00567	0.03020	-3.50	-0.4733	-0.00584	0.00045	-1.61	-0.6505	-0.3485	-0.00403	0.03047	504.12	893.49
o-xylene	-0.9	-3.54	-0.0103	-0.00593	0.03207	-3.62	-0.5318	-0.00609	0.00049	-1.41	-0.6269	-0.3243	-0.00420	0.03238	562.14	1036.56
m-xylene	-0.82	-3.51	0.0010	-0.00580	0.03256	-3.61	-0.5666	-0.00596	0.00042	-1.37	-0.5790	-0.2954	-0.00415	0.03291	572.7	1041.8
p-xylene	-0.81	-3.24	0.0036	-0.00573	0.03269	-3.34	-0.5544	-0.00588	0.00052	-1.34	-0.5631	-0.2870	-0.00412	0.03304	571.68	1039.87
ethyl-benzene	-0.7	-3.79	-0.0089	-0.00574	0.03272	-3.88	-0.5571	-0.00593	0.00043	-1.54	-0.5832	-0.2996	-0.00414	0.03300	569.18	1044.13
tert-butyl-benzene	-0.44	-3.07	0.0107	-0.00600	0.03590	-3.22	-0.6922	-0.00622	0.00063	-0.81	-0.5220	-0.2565	-0.00445	0.03620	667.07	1326.92
biphenyl	-2.64	-6.90	-0.0837	-0.00903	0.03482	-7.12	-0.7587	-0.00955	0.00013	-3.85	-1.1426	-0.6450	-0.00631	0.03518	715.56	1376.95
diphenyl-methane	-2.81	-7.22	-0.0647	-0.00986	0.03665	-7.41	-0.7776	-0.01029	0.00048	-3.98	-1.2111	-0.6802	-0.00695	0.03701	783.05	1543.71
naphtaline	-2.4	-1.91	-0.0654	-0.00787	0.03211	-5.27	-0.6669	-0.00834	0.00018	-0.88	-0.9830	-0.5602	-0.00545	0.03244	600.23	1125.66
anthracene	-3.46	-7.50	-0.0757	-0.01014	0.03615	-7.82	-0.9076	-0.01078	0.00005	-4.65	-1.2439	-0.7061	-0.00693	0.03658	768.96	1521.38
styrene	-1.27	-4.36	-0.0613	-0.00716	0.03066	-4.51	-0.5801	-0.00752	0.00033	-2.42	-0.9129	-0.5205	-0.00505	0.03092	553.32	998.92
phenol	-6.57	-7.09	-0.1767	-0.01271	0.02658	-7.21	-0.6170	-0.01332	0.00007	-6.47	-2.0451	-1.4479	-0.00887	0.02668	469.06	815.9
2-methyl-phenol	-5.87	-6.46	-0.1468	-0.01211	0.02917	-6.62	-0.7125	-0.01274	0.00026	-6.11	-1.7681	-1.2549	-0.00874	0.02932	531.83	961.17
4-methyl-phenol	-6.13	-7.30	-0.1374	-0.01278	0.02911	-7.43	-0.6830	-0.01334	0.00017	-6.28	-2.3967	-1.4909	-0.00795	0.02930	539.04	967.45
4-tert-buthyl-phenol	-5.92	-7.02	-0.1064	-0.01304	0.03489	-7.30	-0.8932	-0.01366	0.00034	-5.45	-2.6160	-1.4238	-0.00756	0.03513	702.4	1400.16
methoxy-benzene	-2.55	-4.76	-0.0697	-0.00915	0.02956	-5.37	-0.5337	-0.00949	0.00027	-3.25	-1.4773	-0.8636	-0.00586	0.02982	535.69	959.17
1-phenyl-ethanone	-4.58	-6.31	0.0106	-0.01333	0.02880	-6.44	-0.4464	-0.01351	-0.00011	-5.26	-1.3786	-1.0182	-0.00975	0.02898	577.54	1061.54
diphenyl-methanone		-6.61	-0.0074	-0.01534	0.03368	-6.84	-0.6762	-0.01574	-0.00028	-4.88	-1.9393	-1.2796	-0.01024	0.03397	790.25	1548
benzaldehyde	-4.02	-6.06	0.0445	-0.01242	0.02638	-6.20	-0.3441	-0.01253	-0.00035	-5.18	-1.3116	-0.9704	-0.00902	0.02657	517.51	923.68
3-hydroxy-benzaldehyde	-9.51	-10.35	-0.1123	-0.01953	0.02572	-10.58	-0.5931	-0.02011	-0.00053	-9.63	-3.9695	-2.4353	-0.01075	0.02594	552.88	998.69
4-hydroxy-benzeldehyde	-10.47	-10.67	-0.1214	-0.02050	0.02574	-10.91	-0.5905	-0.02108	-0.00054	-10.01	-4.4633	-2.5969	-0.01055	0.02598	551.44	996.95

benzoic acid		-7.76	-0.1600	-0.01450	0.02595	-7.93	-0.5697	-0.01502	-0.00034	-7.06	-3.3335	-1.8831	-0.00791	0.02620	546.82	986.46
benzoic acid methyl ester	-4.28	-7.49	-0.0033	-0.01205	0.02888	-7.65	-0.5082	-0.01233	-0.00007	-5.97	-1.5554	-1.0580	-0.00850	0.02912	621.67	1137.38
furane	-0.9	-4.04	-0.0734	-0.00635	0.02216	-4.02	-0.3176	-0.00660	0.00023	-3.18	-0.9888	-0.6384	-0.00447	0.02227	364.43	591.68
carbonmonoxide	2.23	-0.91	-0.0120	-0.00119	0.00964	-0.88	-0.0481	-0.00125	-0.00036	-0.93	-0.0124	-0.0124	-0.00122	0.00964	206.68	271.85
carbondioxide	0.11	-2.38	-0.0743	-0.00388	0.01004	-2.43	-0.2000	-0.00418	-0.00053	-2.43	-0.0762	-0.0762	-0.00397	0.01004	236.59	322.08
hydrogen	2.34	-0.23	0.0081	-0.00035	0.02072	-0.33	-0.1944	-0.00053	0.00023	-0.24	0.0082	0.0082	-0.00035	0.02072	97.76	88.42
water	-6.29	-8.40	-0.2628	-0.01448	0.01221	-8.43	-0.4280	-0.01513	0.00002	-8.59	-0.2707	-0.2707	-0.01484	0.01221	153.91	172.72
hydrogenperoxide		-9.00	-0.3048	-0.01474	0.01204	-8.93	-0.4076	-0.01539	-0.00047	-9.19	-0.3130	-0.3130	-0.01510	0.01204	207.28	261.69
nitrogen	2.45	-0.64	0.0056	-0.00100	0.00953	-0.62	0.0028	-0.00098	-0.00051	-0.04	-0.0476	-0.0208	-0.00087	0.00950	195.52	249.51
ammonia		-6.46	-0.2848	-0.01105	0.01716	-6.60	-0.6375	-0.01194	0.00011	-6.60	-0.2938	-0.2938	-0.01133	0.01716	174.84	207.49
N-methyl-amine		-4.98	-0.1978	-0.00883	0.02181	-5.15	-0.5420	-0.00948	0.00008	-4.99	1.4043	1.7408	-0.01253	0.02180	258.38	357.72
N,N-dimethyl-amine		-3.60	-0.0947	-0.00678	0.02630	-3.81	-0.5604	-0.00724	0.00023	-3.41	0.6696	0.7426	-0.00853	0.02627	332.33	509.42
N,N,N-trimethyl-amine		-3.26	0.0100	-0.00484	0.03001	-3.32	-0.5717	-0.00508	0.00049	-2.86	0.8114	0.9136	-0.00651	0.02993	398.67	663.96
N-ethyl-amine		-4.95	-0.1675	-0.00890	0.02624	-5.16	-0.6615	-0.00959	0.00015	-4.87	1.5956	1.9969	-0.01295	0.02627	330.89	505.99
N,N-diethyl-amine		-3.72	-0.0236	-0.00632	0.03349	-3.90	-0.7244	-0.00675	0.00035	-3.32	0.8519	0.9389	-0.00823	0.03345	476.82	801.14
N,N,N-triethyl-amine		-4.07	0.0507	-0.00360	0.03837	-3.86	-0.8086	-0.00381	0.00132	-5.27	1.5200	2.3454	-0.00679	0.03817	576.18	1081.77
N-propyl-amine		-5.01	-0.1502	-0.00897	0.02944	-5.19	-0.7331	-0.00966	0.00019	-4.83	1.3076	1.5567	-0.01231	0.02948	402.45	654.7
N-iso-propyl-amine		-4.54	-0.1148	-0.00839	0.02927	-4.76	-0.7304	-0.00900	0.00046	-4.30	1.0949	1.2672	-0.01122	0.02924	396.54	655.85
N-butyl-amine		-5.01	-0.1357	-0.00912	0.03238	-5.19	-0.7999	-0.00980	0.00021	-4.74	1.2551	1.4643	-0.01227	0.03243	473.55	802.37
N-tert-butyl-amine		-4.61	-0.0975	-0.00844	0.03114	-4.75	-0.7819	-0.00905	0.00076	-4.31	1.2883	1.5199	-0.01162	0.03113	452.31	809.49
N-pentyl-amine		-4.97	-0.1254	-0.00926	0.03517	-5.17	-0.8799	-0.00994	0.00026	-4.58	1.1632	1.3228	-0.01215	0.03522	544.35	949.66
N-hexyl-amine		-5.01	-0.1185	-0.00937	0.03779	-5.23	-0.9661	-0.01006	0.00027	-4.50	1.0296	1.1334	-0.01191	0.03784	615.29	1096.79
N,N-dipropyl-amine		-4.65	-0.0105	-0.00644	0.03886	-4.76	-0.9386	-0.00696	0.00046	-3.90	0.6180	0.6203	-0.00769	0.03888	619.54	1095.96
N,N-dibutyl-amine		-5.18	0.0041	-0.00685	0.04383	-5.23	-1.0803	-0.00736	0.00042	-4.24	0.7771	0.7835	-0.00837	0.04386	760.73	1388.95
aziridine		-5.92	-0.1220	-0.01011	0.02047	-5.89	-0.3399	-0.01047	0.00010	-5.89	0.9546	1.0799	-0.01280	0.02044	299.74	447.73
pyrrolidine		-4.87	-0.1094	-0.00798	0.02946	-4.97	-0.7018	-0.00861	0.00016	-4.79	1.4424	1.7706	-0.01161	0.02944	415	706.49
piperidine		-3.63	-0.0595	-0.00628	0.03289	-3.76	-0.7268	-0.00671	0.00013	-3.18	0.4462	0.4517	-0.00737	0.03290	460.4	822.77
N-methyl-piperidine		-2.87	0.0357	-0.00455	0.03599	-2.94	-0.7401	-0.00481	0.00029	-2.11	0.6470	0.6600	-0.00571	0.03599	525.59	975.97
piperazine		-9.04	-0.2606	-0.01153	0.03130	-9.08	-1.0996	-0.01269	-0.00016	-9.20	1.0342	1.1684	-0.01449	0.03122	448.99	789.69
ethyl-1,2-diamine		-7.38	-0.2684	-0.01336	0.02716	-7.69	-0.8308	-0.01433	0.00012	-7.27	1.5447	1.8133	-0.01772	0.02710	378.02	603.89

formamide		-10.86	-0.1782	-0.02073	0.01777	-10.85	-0.4102	-0.02124	-0.00017	-11.40	-1.1239	-1.0557	-0.01913	0.01781	274.76	389.23
acetamide		-11.28	-0.1975	-0.02090	0.02138	-11.27	-0.5308	-0.02150	-0.00013	-11.41	-1.8013	-1.5650	-0.01751	0.02145	345.4	533.76
N-methyl-formamid		-10.24	-0.0440	-0.01932	0.02114	-10.21	-0.3593	-0.01965	-0.00003	-10.40	-1.2180	-1.0909	-0.01683	0.02122	347.93	541.68
N-methyl-acetamid	-10.08	-10.79	-0.0540	-0.01945	0.02445	-10.76	-0.4741	-0.01987	-0.00005	-10.86	-0.8856	-0.8162	-0.01782	0.02450	417.87	683.41
N,N-dimethyl-formamid		-9.14	0.0688	-0.01612	0.02415	-8.99	-0.2555	-0.01613	-0.00007	-8.94	0.9477	1.0025	-0.01869	0.02408	414.89	688.88
N,N-dimethyl-acetamid	-8.55	-9.85	0.0460	-0.01632	0.02676	-9.67	-0.3432	-0.01637	0.00003	-9.59	1.1072	1.1810	-0.01934	0.02669	474.22	825.66
carbamid acid methyl ester		-9.50	-0.2255	-0.01741	0.02056	-9.62	-0.5254	-0.01802	-0.00011	-9.60	-2.2041	-1.7771	-0.01350	0.02065	385.26	611.64
aniline	-5.68	-7.26	-0.2568	-0.01312	0.02889	-7.51	-0.8574	-0.01406	0.00032	-6.09	-2.2701	-1.4846	-0.00870	0.02902	484.03	845.85
benzylamine		-6.03	-0.1505	-0.01111	0.03125	-6.27	-0.7473	-0.01177	0.00039	-4.89	-0.8667	-0.6884	-0.00946	0.03139	550.57	990.98
pyrrole	-4.3	-6.83	-0.2948	-0.01203	0.02539	-7.02	-0.7447	-0.01298	0.00054	-6.32	-1.0177	-0.8739	-0.01053	0.02538	378.01	616.17
3-methyl-1H-indole	-5.91	-8.21	-0.2189	-0.01365	0.03240	-8.37	-0.9071	-0.01451	0.00042	-6.49	-1.2776	-0.9794	-0.01097	0.03255	615.79	1156.19
imidazole		-10.73	-0.3729	-0.02048	0.02238	-10.89	-0.7341	-0.02161	0.00010	-10.28	-3.4657	-2.4421	-0.01283	0.02240	363.06	583.92
4-methyl-1H-imidazole	-10.24	-10.07	-0.3310	-0.01951	0.02573	-10.26	-0.7988	-0.02058	0.00015	-9.09	-3.7662	-2.3487	-0.01076	0.02592	434.4	728.52
pyridine	-4.7	-5.52	-0.1205	-0.01034	0.02525	-5.61	-0.4486	-0.01080	-0.00009	-5.07	0.7873	0.8515	-0.01267	0.02521	419.6	711.2
o-picoline	-4.63	-5.26	-0.0917	-0.00992	0.02836	-5.40	-0.5675	-0.01041	0.00012	-4.58	0.7240	0.7612	-0.01187	0.02833	487.88	855.21
m-picoline	-4.77	-5.82	-0.0893	-0.01069	0.02810	-5.93	-0.5166	-0.01112	0.00001	-5.20	0.8408	0.8954	-0.01299	0.02808	487.74	859.05
p-picoline	-4.93	-6.39	-0.0936	-0.01099	0.02813	-6.48	-0.5384	-0.01144	0.00003	-5.81	0.9845	1.0676	-0.01369	0.02811	489.2	861.39
2,6-dimethyl-pyridine	-4.6	-5.04	-0.0648	-0.00956	0.03112	-5.28	-0.7530	-0.01020	0.00030	-4.16	0.6565	0.6692	-0.01116	0.03110	557.78	999.35
chinoline	-5.69	-6.45	-0.1420	-0.01195	0.03050	-6.72	-0.7428	-0.01268	-0.00012	-5.42	-0.1319	-0.1229	-0.01179	0.03056	588.56	1099.15
pyridazine		-11.15	-0.2374	-0.01943	0.02274	-11.17	-0.5618	-0.02029	-0.00078	-11.04	3.3897	4.1806	-0.03027	0.02271	405.61	680.84
pyrimidine		-6.87	-0.2047	-0.01245	0.02290	-7.01	-0.5433	-0.01318	-0.00049	-6.46	1.1252	1.2606	-0.01579	0.02286	401.62	675.22
pyrazine		-7.16	-0.1622	-0.01117	0.02292	-7.17	-0.4704	-0.01175	-0.00049	-6.72	0.8158	0.8843	-0.01346	0.02289	402.63	677.31
2-methyl-pyrazine	-5.52	-6.20	-0.1252	-0.01112	0.02628	-6.38	-0.5629	-0.01170	-0.00026	-5.46	0.8836	0.9455	-0.01339	0.02625	470.58	819.67
hydrogencyanide		-6.42	-0.2055	-0.01214	0.01253	-6.44	-0.3048	-0.01272	-0.00010	-6.22	-0.4598	-0.4349	-0.01169	0.01253	220.6	296.45
acetonitrile	-3.89	-7.36	-0.1042	-0.01482	0.01805	-7.29	-0.2093	-0.01508	-0.00055	-7.19	0.7460	0.7974	-0.01733	0.01805	295.16	430.41
propionitrile	-3.85	-6.98	-0.0255	-0.01399	0.02243	-7.00	-0.2424	-0.01417	-0.00053	-6.35	0.1162	0.1124	-0.01440	0.02244	365.44	572.33
butyronitrile	-3.64	-6.83	0.0316	-0.01380	0.02625	-6.87	-0.2966	-0.01393	-0.00040	-5.85	-0.1851	-0.1685	-0.01313	0.02624	437.66	722.59
isobutyronitrile		-6.77	0.0262	-0.01334	0.02558	-6.80	-0.2903	-0.01345	-0.00028	-5.75	-0.1766	-0.1602	-0.01269	0.02559	433.7	726.43
acrylonitrile	-3.16	-7.14	-0.0889	-0.01311	0.01963	-7.14	-0.2329	-0.01337	-0.00046	-6.57	-0.3565	-0.3294	-0.01245	0.01964	350.61	539.21
benzonitrile	-3.64	-7.33	-0.0555	-0.01389	0.02608	-7.42	-0.3986	-0.01422	-0.00062	-5.57	-0.9048	-0.7352	-0.01153	0.02613	522.29	919.7

3-hydroxy-benzonitrile	-9.66	-10.57	-0.2028	-0.02017	0.02550	-10.80	-0.6393	-0.02095	-0.00077	-9.80	-3.2942	-2.2851	-0.01329	0.02551	555.62	992.43
4-hydroxy-benzonitrile	-10.17	-11.35	-0.2288	-0.02225	0.02547	-11.58	-0.6624	-0.02309	-0.00079	-10.62	-3.7097	-2.5511	-0.01413	0.02548	556.27	992.8
malononitrile		-10.98	-0.2591	-0.02291	0.01817	-11.03	-0.4560	-0.02374	-0.00135	-10.76	0.0891	0.0878	-0.02430	0.01818	386.98	608.04
nitromethane	-4.02	-8.38	-0.0022	-0.01394	0.01589	-8.18	-0.0842	-0.01392	-0.00086	-8.81	-0.9609	-0.8736	-0.01226	0.01592	315.53	471.88
nitroethane	-3.71	-6.89	0.0956	-0.01291	0.01970	-6.91	-0.0785	-0.01275	-0.00081	-6.63	-1.5745	-1.1975	-0.00940	0.01978	384.89	621.99
nitropropane	-3.34	-6.75	0.1573	-0.01268	0.02350	-6.80	-0.1159	-0.01246	-0.00073	-6.07	-1.7930	-1.2336	-0.00851	0.02356	457.3	771.68
nitro-iso-propane	-3.14	-6.16	0.1617	-0.01199	0.02285	-6.20	-0.0987	-0.01172	-0.00050	-5.46	-1.7257	-1.1587	-0.00781	0.02298	445.55	769.52
nitro-benzene	-4.12	-10.60	0.0685	-0.01325	0.02398	-10.57	-0.2121	-0.01321	-0.00072	-8.88	-2.0756	-1.3467	-0.00829	0.02416	533.96	953.5
1-methyl-2-nitro-benzene	-3.59	-7.59	0.1031	-0.01290	0.02654	-7.67	-0.2630	-0.01284	-0.00055	-6.87	-1.9988	-1.2431	-0.00795	0.02675	586.42	1091.59
1-methyl-3-nitro-benzene	-3.45	-6.59	0.1298	-0.01323	0.02669	-6.69	-0.2523	-0.01311	-0.00064	-5.36	-2.1193	-1.2851	-0.00797	0.02687	600.16	1098.11
3-nitro-phenol	-9.63	-10.12	-0.0892	-0.01902	0.02347	-10.28	-0.4560	-0.01944	-0.00089	-10.33	-2.8641	-2.1302	-0.01346	0.02353	567.46	1026.05
4-nitro-phenol	-10.65	-11.59	-0.0964	-0.02227	0.02336	-11.76	-0.4550	-0.02271	-0.00092	-11.53	-3.4253	-2.5055	-0.01493	0.02345	567.94	1026.99
2-nitro-benzonitrile		-12.01	-0.0017	-0.02328	0.02366	-12.14	-0.3086	-0.02352	-0.00140	-11.28	-1.9049	-1.5741	-0.01844	0.02371	606.92	1123.04
3-nitro-benzonitrile		-10.23	-0.0669	-0.02030	0.02380	-10.34	-0.3437	-0.02064	-0.00146	-9.60	-1.8777	-1.5198	-0.01604	0.02379	620.39	1130.37
4-nitro-benzonitrile		-9.77	-0.0694	-0.01957	0.02381	-9.89	-0.3379	-0.01991	-0.00144	-9.04	-1.8657	-1.5001	-0.01550	0.02379	621.6	1132.14
morpholine		-8.82	-0.0938	-0.01168	0.02912	-8.64	-0.6855	-0.01223	-0.00018	-8.74	0.4505	0.4643	-0.01306	0.02909	439.32	766.27
N-methyl-morpholine		-7.69	0.0225	-0.00965	0.03249	-7.47	-0.6888	-0.01000	-0.00001	-7.04	0.4545	0.4513	-0.01059	0.03249	504.79	919.78
chloromethane	-0.55	-4.21	-0.0908	-0.00530	0.01954	-4.16	-0.2742	-0.00567	-0.00057	-2.51	-0.5787	-0.3678	-0.00404	0.01957	420.39	714.39
dichloromethane	-1.18	-4.46	-0.1019	-0.00715	0.01872	-4.39	-0.2340	-0.00744	-0.00079	-3.31	-0.8633	-0.5964	-0.00527	0.01879	351.75	558.23
trichloromethane	-1.05	-3.45	-0.0469	-0.00574	0.01818	-3.35	-0.1416	-0.00580	-0.00039	-2.51	-0.7566	-0.4925	-0.00401	0.01827	279.06	406.35
tetrachloromethane	0.1	-3.35	-0.0623	-0.00169	0.02022	-3.24	-0.2739	-0.00205	0.00012	-3.41	-0.0636	-0.0636	-0.00172	0.02022	479.38	866.83
chloroethane	-0.46	-3.58	0.0077	-0.00606	0.02266	-3.49	-0.2320	-0.00612	-0.00035	-2.25	-0.7608	-0.4508	-0.00401	0.02281	350.55	551.52
1,1-dichloro-ethane	-0.86	-3.83	-0.0253	-0.00706	0.02224	-3.99	-0.3391	-0.00736	-0.00042	-2.45	-0.9265	-0.5545	-0.00473	0.02237	416.08	705.47
hexanchloro-ethane	-1.06	-5.07	-0.0667	-0.00246	0.02196	-4.91	-0.4712	-0.00310	0.00093	-0.87	0.1198	0.0494	-0.00268	0.02198	628.85	1299.64
1-chloro-propane	-0.35	-3.13	0.0495	-0.00587	0.02628	-3.22	-0.2985	-0.00593	-0.00021	-1.62	-0.6891	-0.3740	-0.00390	0.02639	421.45	701.16
2-chloro-propane	-0.25	-3.35	0.0425	-0.00637	0.02590	-3.48	-0.3708	-0.00651	0.00001	-1.85	-0.7516	-0.4194	-0.00412	0.02607	414.36	701.65
1-chloro-butane	-0.14	-3.16	0.0700	-0.00599	0.02944	-3.27	-0.3873	-0.00606	-0.00015	-1.36	-0.6340	-0.3277	-0.00402	0.02953	493.37	850.99
1-chloro-pentane	-0.07	-3.19	0.0853	-0.00611	0.03234	-3.31	-0.4707	-0.00619	-0.00012	-1.13	-0.5766	-0.2876	-0.00418	0.03243	563.19	997.53

2-chloro-pentane	0.07	-3.22	0.0851	-0.00630	0.03175	-3.39	-0.5288	-0.00644	0.00019	-1.17	-0.5765	-0.2894	-0.00413	0.03192	548.15	993.25
1,1-dichloro-ethene	-0.22	-3.83	-0.0253	-0.00706	0.02224	-3.99	-0.3391	-0.00736	-0.00042	-2.45	-0.9265	-0.5545	-0.00473	0.02237	416.08	705.47
cis-1,2-dichloro-ethene	-1.17	-4.13	-0.0471	-0.00612	0.02035	-4.07	-0.2550	-0.00635	-0.00035	-2.55	-0.7342	-0.4448	-0.00421	0.02050	397.9	656.83
trans-1,2-dichloro-ethene	0.76	-3.25	-0.0747	-0.00447	0.02030	-3.17	-0.2394	-0.00471	-0.00060	-1.66	-0.5404	-0.3086	-0.00341	0.02035	404.78	662.61
trichloro-ethene	-0.5	-3.07	-0.0560	-0.00356	0.02050	-3.05	-0.2496	-0.00384	-0.00034	-1.00	-0.3068	-0.1486	-0.00276	0.02057	467.81	811.86
tetrachloro-ethene	-0.16	-0.88	-0.0487	-0.00180	0.02090	-1.03	-0.2699	-0.00211	-0.00004	0.77	0.1871	0.0743	-0.00216	0.02088	526.72	955.57
chloro-benzene	-1.12	-4.56	-0.0297	-0.00619	0.02628	-4.65	-0.4042	-0.00646	-0.00014	-2.50	-0.7503	-0.4119	-0.00432	0.02649	503.83	895.64
1,2-dichloro-benzene	-1.56	-5.91	-0.0124	-0.00665	0.02562	-6.01	-0.4338	-0.00699	-0.00012	-3.10	-0.7482	-0.3920	-0.00445	0.02586	562.05	1039.47
1,3-dichloro-benzene	-1.25	-3.16	-0.0517	-0.00598	0.02556	-3.37	-0.4528	-0.00639	-0.00053	-1.23	-0.6752	-0.3519	-0.00429	0.02573	572.42	1048.4
1,4-dichloro-benzene	-1.38	-3.72	-0.0576	-0.00623	0.02551	-3.88	-0.4633	-0.00667	-0.00054	-1.67	-0.7317	-0.3905	-0.00451	0.02565	573.24	1050.68
hexam-chloro-benzene	-1.78	-2.12	-0.0434	-0.00347	0.02497	-2.48	-0.5566	-0.00410	0.00023	0.62	0.1086	0.0447	-0.00356	0.02494	789.53	1612.65